STATISTICAL METHODS FOR BEHAVIORAL SCIENCE RESEARCH

McGRAW-HILL SERIES IN PSYCHOLOGY

CONSULTING EDITORS Norman Garmezy, Richard L. Solomon, Lyle V. Jones, Harold W. Stevenson

ADAMS Human Memory
BEACH, HEBB, MORGAN, AND NISSEN The Neuropsychology of Lashley
VON BÉKÉSY Experiments in Hearing
BERKOWITZ Aggression: A Social Psychological Analysis
BERLYNE Conflict, Arousal, and Curiosity
BLUM Psychoanalytic Theories of Personality
BROWN The Motivation of Behavior
BROWN AND GHISELLI Scientific Method in Psychology
BUTCHER MMPI Research Developments and Clinical Applications
CAMPBELL, DUNNETTE, LAWLER, AND WEICK Managerial Effectiveness
COFER Verbal Learning and Verbal Behavior
COFER AND MUSGRAVE Verbal Behavior and Learning Problems and Processes
CRAFTS, SCHNEIRLA, ROBINSON, AND GILBERT Recent Experiments in Psychology
CRITES Vocational Psychology
D'AMATO Experimental Psychology Methodology, Psychophysics, and Learning
DAVITZ The Communication of Emotional Meaning
DEESE AND HULSE The Psychology of Learning
DOLLARD AND MILLER Personality and Psychotherapy
EDGINGTON Statistical Inference. The Distribution-free Approach
ELLIS Handbook of Mental Deficiency
EPSTEIN Varieties of Perceptual Learning
FERGUSON Statistical Analysis in Psychology and Education
FORGUS Perception The Basic Process in Cognitive Development
FRANKS Behavioral Therapy. Appraisal and Status
GHISELLI Theory of Psychological Measurement
GHISELLI AND BROWN Personnel and Industrial Psychology
GILMER Industrial Psychology
GILMER Industrial and Organizational Psychology
GRAY Psychology Applied to Human Affairs
GUILFORD Fundamental Statistics in Psychology and Education
GUILFORD The Nature of Human Intelligence
GUILFORD Personality
GUILFORD Psychometric Methods
GUILFORD AND HOEPFNER The Analysis of Intelligence
GUION Personnel Testing
HAIRE Psychology in Management
HIRSCH Behavior-genetic Analysis
HIRSH The Measurement of Hearing
HURLOCK Adolescent Development
HURLOCK Child Development
HURLOCK Developmental Psychology

JACKSON AND MESSICK Problems in Human Assessment
KARN AND GILMER Readings in Industrial and Business Psychology
KRECH, CRUTCHFIELD, AND BALLACHEY Individual in Society
LAWLER Pay and Organizational Effectiveness A Psychological View
LAZARUS, A. Behavior Therapy and Beyond
LAZARUS, R. Adjustment and Personality
LAZARUS, R. Psychological Stress and the Coping Process
LEWIN A Dynamic Theory of Personality
LEWIN Principles of Topological Psychology
MAHER Principles of Psychopathology
MARASCUILO Statistical Methods for Behavioral Science Research
MARX AND HILLIX Systems and Theories in Psychology
MESSICK AND BRAYFIELD Decision and Choice: Contributions of Sidney Siegel
MILLER Language and Communication
MORGAN Physiological Psychology
NUNNALLY Psychometric Theory
RETHLINGSHAFER Motivation as Related to Personality
ROBINSON AND ROBINSON The Mentally Retarded Child
ROSENTHAL Genetic Theory and Abnormal Behavior
SCHERER AND WERTHEIMER A Psycholinguistic Experiment on Foreign Language Teaching
SHAW Group Dynamics The Psychology of Small Group Behavior
SHAW AND WRIGHT Scales for the Measurement of Attitudes
SIDOWSKI Experimental Methods and Instrumentation in Psychology
SIEGEL Nonparametric Statistics for the Behavioral Sciences
SPENCER AND KASS Perspectives in Child Psychology
STAGNER Psychology of Personality
TOWNSEND Introduction to Experimental Methods for Psychology and the Social Sciences
VINACKE The Psychology of Thinking
WALLEN Clinical Psychology The Study of Persons
WARREN AND AKERT The Frontal Granular Cortex and Behavior
WATERS, RETHLINGSHAFER, AND CALDWELL Principles of Comparative Psychology
WINER Statistical Principles in Experimental Design
ZUBEK AND SOLBERG Human Development

John F. Dashiell was Consulting Editor of this series from its inception in 1931 until January 1, 1950. Clifford T. Morgan was Consulting Editor of this series from January 1, 1950 until January 1, 1959. Harry F. Harlow assumed the duties of Consulting Editor from 1959 to 1965. In 1965 a Board of Consulting Editors was established according to areas of interest. The current board members are Richard L. Solomon (physiological, experimental), Norman Garmezy (abnormal, clinical), Harold W. Stevenson (child, adolescent, human development), and Lyle V. Jones (statistical, quantitative).

This book was set in Monotype Times Roman by William Clowes & Sons Ltd , and printed on permanent paper and bound by The Maple Press Company. The designer was Janet Bollow; the drawings were done by Judith McCarty. The editors were Walter J. Maytham and Michael A. Ungersma. Charles A. Goehring supervised production.

STATISTICAL METHODS FOR BEHAVIORAL SCIENCE RESEARCH

Printed in the United States of America.

Library of Congress catalog card number 70-132347

1234567890 MAMM 7987654321

07–040185–3

TO
NICHOLAS
AND
DORABELLA

PREFACE

This book presents, in as nonmathematical a manner as possible, the basic philosophy and theory of modern-day statistics for students and researchers in the behavioral sciences. The primary reason for emphasizing theory is that many contemporary behavioral researchers use the techniques described here in a mechanical manner, often with little or no understanding of their theoretical basis or their actual limitations Because of this nontheoretical understanding of basic statistical methods, many researchers are at a loss when faced with a research question that cannot be cast into one of the molds or statistical models that were learned in a traditional one- or two-semester course in statistics and research design. It is my belief that novice researchers should not face this predicament, since statistical theory is quite simple, easy to understand, and often based on ordinary common sense.

Another reason why I have emphasized theory is that new and more efficient statistical techniques are appearing in the statistical literature at an increasing rate. Researchers who have a poor or minimal understanding of the underlying statistical theory will experience a low probability of success with these new methods of analysing data Part of this subjective prediction is based on the assumption that the learning of techniques, to the exclusion of theory, leads one to believe that some of the new and advanced techniques are more difficult to use and understand, since higher mathematics and arithmetic skills are required. Even though the statistical theory is simple, the mathematics often gives a false aura of complexity to advanced methods. To help eliminate this problem, I have attempted to present statistical theory in a manner that will assist researchers to prepare for the more elaborate procedures that are now available in the technical literature, or that will appear in the applied literature in the near future.

When I first began to teach elementary statistics, I was advised to teach students about the mean, median, mode, standard deviation, correlation coefficient, and the *t* test. I did this with what I imagined to be some success. Later, when students began to consult about their dissertation proposals, I quickly learned that what I had taught was of limited value to them in the solution of their problems. Their research question invariably required knowledge of main effects, interactions, crossed factors, nested factors, *post hoc* comparisons, planned compari-

sons, factor analysis, multiple regression, linear discriminant analysis, canonical analysis, and even multivariate analysis of variance, none of which are discussed in this book. What I had taught was basic descriptive statistics, but what they needed was a firm understanding of inferential statistics.

As a result, I modified the course and began to introduce basic theory of probability and statistics. The theory of confidence intervals was added, and the theory of hypothesis testing soon followed. The course then expanded to a two-quarter course in which analysis of variance, contingency table analysis, and correlation and regression were added.

Most of the students I teach have had little or no training in mathematics. I have tried to minimize the algebra by placing it in proofs, which can be ignored by students who are frightened by algebraic manipulations or who do not understand or appreciate a mathematical proof. Unfortunately, statistics by nature is mathematical, and so it is not possible to teach statistics or write a statistics book without mathematics. However, the principles of statistics can be understood even by students who are afraid of algebra Once the principles are understood, the probability that a student can use the principles in a more complex and elaborate situation is quite high. While I do not have any statistics to defend this statement, I have seen dozens of students who have come with fear and hostility into my classes work hard, put in extra hours of studying, and finally emerge more confident of their use of statistics, and, what is more important, I have actually seen them use their knowledge in more advanced and complex statistical methods with considerable success. I have seen the payoff to this approach. It is real!

Even though this is a spiral presentation of material built on previously discovered concepts, the coverage is broad enough so that each individual instructor can pick and choose to satisfy his own needs Without doubt, supplementary readings will be required and the teaching of specialized skills will have to be considered. For example, use of desk calculators and the use of IBM cards and computers may be arranged. Students may be instructed in the use of canned programs for data description, analysis of variance, contingency-table analysis, and correlation and regression. The class also may do a random sampling study, using the list of 500 scores reported in Table 8-1. In addition, the class may compare their group results to those reported in Chapter 9 for one class that also studied sampling distributions. Sampling distributions and power have always been among the most difficult statistical concepts to teach. In addition, they appear to be the most difficult for the students to comprehend. Particular effort has been made to make the explanations of these topics clear.

Without doubt, the first twelve chapters of this text are the most complex and most difficult for most students. Once these chapters are mastered, the rest of the book is much easier. Thus, the beginning student is urged to pay extra attention to this first half. Exercises are included following each chapter, to help the student gain an understanding of the principles discussed. These problems range from very easy to rather difficult. Most of them are modifications of real problems brought to me by faculty and students. Some of the problems have been starred. This does not mean that these problems are difficult, instead, it means that the results will be used in some problem or problems that will appear later in the book. Thus, some of the problems are used to create a continuity between certain concepts that are presented in the various chapters. For these and some of the other problems, an answer key is provided at the end of the book. A listing of related readings is also presented for those students who wish to study a particular set of concepts in greater detail.

As would be expected, a beginning book in elementary statistics requires the assistance of many people. It was Professor Malcolm Slakter who first induced me to begin writing this text, Dr. Seong Soo Lee, Dr. Maryellen McSweeney, Dr. Douglas Penfield, Dr. Joel Levin, and

Dr. Neil Timm have extensively reviewed and tested the material in classes. Many of their recommendations, and those of their students, have made their way into this text. The manuscript and exercises were checked and solved by Mr. Darshan Sacdeva, Mr Steve Lawton, and Mr John Hart. The typing of the text and its many revisions was done by Mr. Thomas Little, with great care and patience. I would like to thank them all—and any others whose names I might have omitted—for their cooperation.

Finally, I would like to thank the various authors and publishers for permission to use the quotations at the head of the chapters, where specific credits appear Special thanks go to the editors of *Technometrics* and Dr. Frank Grubbs for their permission to reproduce Table A-6, and to W. J. Dixon and F. J. Massey for their permission to utilize in Appendix A a number of tables from *Introduction to Statistical Analysis.* I am indebted to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Oliver & Boyd, Edinburgh, for permission to reprint Table A-8 from their book *Statistical Tables for Biological, Agricultural and Medical Research.*

LEONARD A. MARASCUILO

CONTENTS

1 INTRODUCTION TO MODERN STATISTICS 1

1-1 WHAT IS STATISTICS? 2
1-2 AN ILLUSTRATION OF THE USE OF MODERN STATISTICAL METHODS 3
1-3 STATISTICS AS A RESEARCH TOOL OF BEHAVIORAL RESEARCH 5
1-4 WHAT STATISTICS IS ABOUT 6
1-5 KINDS OF STATISTICAL VARIABLES 9
1-6 SUMMARY 13
EXERCISES 14

2 INTRODUCTION TO SET THEORY IN BEHAVIORAL RESEARCH 16

2-1 POPULATIONS 17
2-2 INTRODUCTION TO SETS 19
2-3 THE EMPTY SET 21
2-4 EQUAL SETS AND SUBSETS 22
2-5 COMPLEMENTARY SETS 22
2-6 THE INTERSECTION OF TWO SETS 23
2-7 MUTUALLY EXCLUSIVE OR DISJOINT SETS 23
2-8 THE UNION OF TWO SETS 23
2-9 VENN DIAGRAMS 24
2-10 SET REPRESENTATION BY TABLES 27
2-11 SETS AS EVENTS 29
2-12 THE NUMBER OF ELEMENTS OF A FINITE SET 29
2-13 THE CARTESIAN PRODUCT OF SETS—SAMPLE SPACES 30
2-14 SUMMARY 32
EXERCISES 33

3 INTRODUCTION TO PROBABILITY THEORY 36

3-1 THE PROBABILITY OF AN EVENT FOR EQUALLY LIKELY OUTCOMES 37
3-2 GENERAL PROBABILITY MODEL 38

3-3 PROBABILITIES OF EVENTS IN FINITE SAMPLE SPACES 40
3-4 PROBABILITIES OF EVENTS IN A TWO-DIMENSIONAL SAMPLE SPACE 41
3-5 THE RELATIVE-FREQUENCY INTERPRETATION OF PROBABILITY 42
3-6 PROBABILITY OF THE COMPLEMENT OF *A* 45
3-7 THE ADDITION RULE FOR MUTUALLY EXCLUSIVE EVENTS 46
3-8 PROBABILITIES OF EVENTS THAT ARE NOT MUTUALLY EXCLUSIVE 47
3-9 STATISTICAL INDEPENDENCE OF TWO EVENTS 49
3-10 CONDITIONAL PROBABILITY 52
3-11 INTRODUCTION TO GENERAL SAMPLING THEORY 58
3-12 SAMPLING WITHOUT REPLACEMENT 59
3-13 SAMPLING WITH REPLACEMENT 61
3-14 SUMMARY 63
EXERCISES 64

4 INTRODUCTION TO COMBINATORIAL ANALYSIS 67

4-1 FUNDAMENTALS OF COUNTING 68
4-2 PERMUTATIONS 69
4-3 COMBINATIONS 75
4-4 THE ALGEBRA OF THE SUMMATION SYMBOL $\sum$ 79
4-5 SUMMARY 81
EXERCISES 82

5 BERNOULLI VARIABLES 84

5-1 COIN TOSSING AS A BEHAVIORAL RESEARCH MODEL 85
5-2 INTRODUCTION TO BERNOULLI AND BINOMIAL VARIABLES 86
5-3 ERRORS IN HYPOTHESIS TESTING 90
5-4 THE BINOMIAL FORMULA 92
5-5 HYPOTHESIS TESTING—AN EXAMPLE 94
5-6 THE SIGN TEST 97
5-7 INTRODUCTION TO SAMPLING THEORY—CONTINUATION 99
5-8 THE HYPERGEOMETRIC FORMULA 102
5-9 THE IRWIN–FISHER EXACT TEST 104
5-10 SUMMARY 105
EXERCISES 107

6 DISCRETE PROBABILITY DISTRIBUTIONS 110

6-1 DISCRETE RANDOM VARIABLES 111
6-2 DISCRETE RANDOM VARIABLES GENERATED FROM OTHER DISCRETE RANDOM VARIABLES 114
6-3 GRAPHIC SUMMARY PROCEDURES FOR DISCRETE RANDOM VARIABLES 117
6-4 THE MEDIAN VALUE OF A DISCRETE RANDOM VARIABLE 119
6-5 THE EXPECTED VALUE OR POPULATION MEAN OF A DISCRETE RANDOM VARIABLE 120
6-6 THE EXPECTED VALUE OF A BINOMIAL RANDOM VARIABLE 124
6-7 LONG-RUN INTERPRETATION OF THE EXPECTED VALUE OF A RANDOM VARIABLE 125
6-8 SKEWED-PROBABILITY DISTRIBUTION 126
6-9 THE MODE OF A PROBABILITY DISTRIBUTION 126
6-10 VARIANCE OF A PROBABILITY DISTRIBUTION 126

6-11 VARIANCE OF A BINOMIAL VARIABLE 131
6-12 THE EXPECTED VALUE AND VARIANCE OF A HYPERGEOMETRIC VARIABLE 132
6-13 THE COMPUTING FORMULA FOR THE VARIANCE OF A DISCRETE RANDOM VARIABLE 133
6-14 PARAMETERS OF PROBABILITY DISTRIBUTIONS 135
6-15 SUMMARY 135
EXERCISES 137

7 CONTINUOUS PROBABILITY DISTRIBUTIONS 140

7-1 CONTINUOUS RANDOM VARIABLES 141
7-2 GRAPHIC METHODS FOR CONTINUOUS VARIABLES 142
7-3 QUARTILES, DECILES, AND PERCENTILE VALUES OF PROBABILITY DISTRIBUTIONS 143
7-4 CONTINUOUS DISTRIBUTIONS DEFINED AS SMOOTH CURVES 144
7-5 THE NORMAL DISTRIBUTION 146
7-6 EXAMPLE OF A NORMALLY DISTRIBUTED RANDOM VARIABLE 149
7-7 THE ALGEBRA OF EXPECTED VALUES 150
7-8 TABLE OF THE NORMAL DISTRIBUTION 156
7-9 CORRECTION OF DISCRETE SCORES TO FIT A CONTINUOUS PROBABILITY DISTRIBUTION 157
7-10 TRANSFORMED Z SCORES AND T SCORES 158
7-11 THE CENTRAL LIMIT THEOREM 162
7-12 THE NORMAL APPROXIMATION TO THE BINOMIAL 167
7-13 PARAMETERS OF PROBABILITY DISTRIBUTIONS (CONTINUATION) 168
7-14 SUMMARY 168
EXERCISES 170

8 INTRODUCTION TO THE STATISTICAL THEORY OF ESTIMATION 173

8-1 RANDOM SAMPLING 174
8-2 SELECTING A SAMPLE FROM A FINITE POPULATION 175
8-3 FREQUENCY TABLES 179
8-4 SAMPLE HISTOGRAMS, RELATIVE-FREQUENCY POLYGONS, AND CUMULATIVE RELATIVE-FREQUENCY POLYGONS 181
8-5 THE MEDIAN AND OTHER PERCENTILES OF A SAMPLE 181
8-6 THE MEAN OR AVERAGE VALUE OF A SAMPLE 185
8-7 COMPUTING FORMULA FOR THE AVERAGE ON A TRANSFORMED SCALE 188
8-8 THE VARIANCE AND STANDARD DEVIATION OF A SAMPLE 190
8-9 COMPUTING FORMULA FOR A SAMPLE VARIANCE ON A TRANSFORMED SCALE 191
8-10 SAMPLE MEAN OR AVERAGE AND SAMPLE MEDIAN AS ESTIMATORS OF THE POPULATION CENTER 193
8-11 SUMMARY 195
EXERCISES 197

9 SAMPLING DISTRIBUTIONS OF STATISTICAL ESTIMATORS 201

9-1 PROPERTIES OF POINT ESTIMATORS 202
9-2 THE SAMPLING DISTRIBUTION OF $\bar{X}$ AND $\hat{M}$ 203

9-3 UNBIASED ESTIMATORS (COMPARISON OF $\bar{X}$ AND $\hat{M}$) 207
9-4 VARIANCES OF SAMPLING DISTRIBUTIONS 209
9-5 EFFICIENT ESTIMATORS 210
9-6 THE VARIANCE OF THE SAMPLING DISTRIBUTION OF $\bar{X}$ 212
9-7 THE THEORETICAL SAMPLING DISTRIBUTION OF MEANS AND MEDIANS 213
9-8 USES OF SAMPLING DISTRIBUTIONS 216
9-9 SAMPLING DISTRIBUTION OF VARIANCE MEASURES 218
9-10 THE SAMPLING DISTRIBUTION OF $\hat{p}$, THE ESTIMATOR OF THE BINOMIAL DISTRIBUTION PARAMETER 223
9-11 THE VARIANCE OF THE SAMPLING DISTRIBUTION OF $\hat{p}$, THE ESTIMATOR OF A HYPERGEOMETRIC DISTRIBUTION PARAMETER 225
9-12 THE VARIANCE OF THE SAMPLING DISTRIBUTION OF $\bar{X}$ FOR FINITE UNIVERSES 225
9-13 SUMMARY 226
EXERCISES 227

10 THE CHI-SQUARE AND t DISTRIBUTIONS 229

10-1 INTRODUCTION TO THE CHI-SQUARE AND t-DISTRIBUTIONS 230
10-2 DEGREES OF FREEDOM ASSOCIATED WITH A SAMPLE VARIANCE 230
10-3 THE CHI-SQUARE DISTRIBUTION 232
10-4 THE SAMPLING DISTRIBUTION OF S^2 (CONTINUATION) 236
10-5 THE t-DISTRIBUTION 239
10-6 SAMPLING DISTRIBUTIONS AND PARAMETER INFERENCES 241
10-7 SUMMARY 242
EXERCISES 244

11 INTERVAL ESTIMATION 245

11-1 INTRODUCTION TO INTERVAL ESTIMATION (PRECISION OF POINT ESTIMATES) 246
11-2 CONFIDENCE INTERVAL FOR μ, ASSUMING σ^2 IS KNOWN 247
11-3 THE LARGE-SAMPLE CONFIDENCE INTERVAL FOR p, THE PARAMETER OF A BINOMIAL DISTRIBUTION 250
11-4 CONFIDENCE INTERVALS FOR μ WHEN σ^2 IS UNKNOWN 252
11-5 THE STANDARD ERROR OF A STATISTIC 253
11-6 CONFIDENCE INTERVAL FOR A GENERAL PARAMETER 254
11-7 CONFIDENCE INTERVAL FOR THE DIFFERENCES OF TWO EXPECTED VALUES (VARIANCES KNOWN) 255
11-8 CONFIDENCE INTERVAL FOR THE DIFFERENCE OF TWO EXPECTED VALUES WHEN VARIANCES ARE UNKNOWN AND EQUAL 257
11-9 CONFIDENCE INTERVALS FOR THE DIFFERENCE OF TWO EXPECTED VALUES WHEN THE VARIANCES ARE UNKNOWN AND NOT NECESSARILY EQUAL 261
11-10 CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO EXPECTED VALUES WHEN THE SAMPLES ARE NOT STATISTICALLY INDEPENDENT (COMMON VARIANCE NOT ASSUMED) 263
11-11 CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO BINOMIAL PARAMETERS 265
11-12 CONFIDENCE INTERVAL FOR σ^2 OF THE NORMAL DISTRIBUTION 267

11-13	INTRODUCTION TO EXPERIMENTAL DESIGN	269
11-14	SUMMARY	271
	EXERCISES	274

12 HYPOTHESIS TESTING IN THE ONE-SAMPLE MODEL 277

12-1	INTRODUCTION TO THE GENERAL THEORY OF HYPOTHESIS TESTING	278
12-2	ONE-SAMPLE TEST FOR EXPECTED VALUES WHEN THE POPULATION VARIANCE IS KNOWN	288
12-3	ONE-SAMPLE TEST FOR EXPECTED VALUES WHEN THE POPULATION VARIANCE IS UNKNOWN	291
12-4	ONE-SAMPLE TEST FOR THE PARAMETER OF A BINOMIALLY DISTRIBUTED VARIABLE	293
12-5	ONE-SAMPLE TEST FOR THE PARAMETER OF THE HYPERGEOMETRIC DISTRIBUTION	296
12-6	ONE-SAMPLE TEST FOR THE VARIANCE OF A POPULATION	297
12-7	DECISION RULES AND SAMPLE SPACES	298
12-8	CONFIDENCE INTERVALS AND HYPOTHESIS TESTING	299
12-9	TEST OF HYPOTHESIS FOR A GENERAL PARAMETER	300
12-10	THE POWER OF THE GENERAL TEST OF HYPOTHESIS IF $\sigma_{\hat{\theta}}^2$ IS KNOWN	301
12-11	H_0 VERSUS H_1 OR H_1 VERSUS H_0?	302
12-12	SUMMARY	303
	EXERCISES	305

13 HYPOTHESIS TESTING IN THE TWO-SAMPLE MODEL 307

13-1	TEST OF HYPOTHESIS FOR THE DIFFERENCE OF TWO EXPECTED VALUES (VARIANCES KNOWN)	308
13-2	TESTS OF HYPOTHESIS FOR THE DIFFERENCE OF TWO EXPECTED VALUES (VARIANCES UNKNOWN BUT ASSUMED TO BE EQUAL)	310
13-3	TEST OF HYPOTHESIS FOR THE DIFFERENCE OF TWO EXPECTED VALUES (VARIANCES UNKNOWN AND NOT NECESSARILY EQUAL)	312
13-4	TEST OF HYPOTHESIS FOR THE DIFFERENCE OF TWO EXPECTED VALUES WHEN THE SAMPLES ARE NOT STATISTICALLY INDEPENDENT	315
13-5	TESTS OF HYPOTHESIS FOR THE DIFFERENCE OF THE PARAMETERS OF TWO BINOMIAL VARIABLES	318
13-6	CONDITIONAL PROBABILITIES AS MEASURES OF RELATIONSHIP	321
13-7	TWO-SAMPLE TESTS AND CONFIDENCE INTERVALS	322
13-8	SUMMARY	322
	EXERCISES	325

14 THE *F* DISTRIBUTION AND TESTS OF VARIANCE 329

14-1	THE *F* DISTRIBUTION	330
14-2	THE *F* TEST FOR THE EQUALITY OF TWO VARIANCES FROM NORMAL POPULATIONS	331
14-3	CONFIDENCE INTERVAL FOR THE RATIO OF TWO VARIANCES OF NORMAL POPULATIONS	333
14-4	ESTIMATION OF THE VARIANCE FROM *K* INDEPENDENT SAMPLES	334
14-5	EXPLAINED VARIABILITY	336
14-6	THE SAMPLING DISTRIBUTION OF Sp^2, THE POOLED ESTIMATE OF σ^2	338

14-7 COCHRAN'S TEST FOR THE EQUALITY OF K-POPULATION VARIANCES 340
14-8 THE RELATIONSHIPS AMONG THE t, F, AND χ^2 DISTRIBUTIONS 341
14-9 TESTS AND CONFIDENCE INTERVALS FOR VARIANCES IN THE TWO- AND K-SAMPLE MODEL 343
14-10 SUMMARY 343
EXERCISES 346

15 HYPOTHESIS TESTING IN THE K-SAMPLE QUANTITATIVE ANALYSIS-OF-VARIANCE MODEL 348

15-1 INTRODUCTORY REMARKS CONCERNING THE K-SAMPLE PROBLEM 349
15-2 THE ONE-WAY ANALYSIS OF VARIANCE—THE K-SAMPLE EXTENSION OF THE TWO-SAMPLE t TEST 349
15-3 COMPUTATION FORMULAS FOR THE SUM OF SQUARES IN AN ANALYSIS-OF-VARIANCE MODEL 354
15-4 THE ANALYSIS-OF-VARIANCE TABLE 356
15-5 SCHEFFÉ'S METHOD OF MULTIPLE CONTRASTS—*POST HOC* COMPARISONS 356
15-6 TUKEY'S METHOD OF MULTIPLE PAIRWISE CONTRASTS 360
15-7 THE F^* TEST THE K-SAMPLE ANALOG TO THE F TEST WHEN POPULATION VARIANCES ARE UNEQUAL 361
15-8 MULTIPLE CONTRASTS FOR THE F^* TEST 364
15-9 EXPLAINED VARIANCE IN THE ANALYSIS-OF-VARIANCE MODEL 364
15-10 TESTS AND CONFIDENCE INTERVALS FOR THE QUANTITATIVE K-SAMPLE MODEL 365
15-11 SUMMARY 367
EXERCISES 370

16 HYPOTHESIS TESTING IN THE K-SAMPLE QUALITATIVE CHI-SQUARE MODEL 373

16-1 THE TWO-SAMPLE BINOMIAL TEST AS A CHI-SQUARE TEST 374
16-2 THE K-SAMPLE BINOMIAL TEST 377
16-3 *POST HOC* PROCEDURES FOR THE K-SAMPLE BINOMIAL PROBLEM 380
16-4 EXTENSION OF THE K-SAMPLE BINOMIAL PROBLEM TO THE K-SAMPLE CHI-SQUARE TEST OF HOMOGENEITY 382
16-5 SIMPLE COMPUTING FORMULA FOR CHI-SQUARE 387
16-6 TABLE OF THE STATISTICAL TESTS AND CONFIDENCE INTERVALS FOR CHI-SQUARE TESTS OF HOMOGENEITY 388
16-7 SUMMARY 388
EXERCISES 392

17 INTRODUCTION TO CORRELATION THEORY IN THE QUALITATIVE MODEL 394

17-1 CHI-SQUARE TESTS OF HOMOGENEITY AND INDEPENDENCE—A COMPARISON 395
17-2 CHI-SQUARE TEST OF INDEPENDENCE FOR A 2×2 CONTINGENCY TABLE 395
17-3 THE CHI-SQUARE TEST OF INDEPENDENCE FOR $I \times J$ CONTINGENCY TABLES 398

17-4 STRENGTH OF ASSOCIATION IN THE QUALITATIVE CORRELATION MODEL 401
17-5 THE CORRELATION COEFFICIENT FOR A 2×2 CONTINGENCY TABLE 404
17-6 THE CORRELATION COEFFICIENT FOR AN $I \times J$ CONTINGENCY TABLE 406
17-7 THE EFFECT OF SAMPLE SIZE ON X^2 AND $\hat{\phi}$ 407
17-8 SUMMARY OF THE STATISTICAL TESTS AND CORRELATION MEASURES FOR CHI-SQUARE TESTS OF INDEPENDENCE 409
17-9 SUMMARY 409
EXERCISES 413

18 INTRODUCTION TO STATISTICAL CORRELATION THEORY 416

18-1 THE BIVARIATE NORMAL DISTRIBUTION 417
18-2 THE GEOMETRY OF THE BIVARIATE NORMAL DISTRIBUTION 417
18-3 ESTIMATORS FOR THE PARAMETERS OF THE BIVARIATE NORMAL DISTRIBUTION 423
18-4 AN INTERPRETATION OF THE SAMPLE CORRELATION COEFFICIENT 425
18-5 EXPLAINED VARIANCE FOR CORRELATIONAL MODELS AS MEASURES OF RELATIONSHIP 429
18-6 COMPUTATIONAL FORMULA FOR r 433
18-7 AN EXAMPLE OF A CORRELATIONAL STUDY 435
18-8 THE MEANING OF A REGRESSION LINE 439
18-9 SUMMARY 440
EXERCISES 442

19 INFERENCE PROCEDURES FOR CORRELATED VARIABLES 445

19-1 THE VARIANCE OF CORRELATED VARIABLES 446
19-2 CONFIDENCE INTERVALS FOR THE PARAMETERS OF THE BIVARIATE NORMAL DISTRIBUTION 449
19-3 TESTS OF HYPOTHESIS FOR CORRELATION COEFFICIENTS 450
19-4 SPEARMAN'S RANK CORRELATION COEFFICIENT 455
19-5 REPEATED MEASURES—THE EXTENSION OF THE MATCH PAIR t TEST 458
19-6 COCHRAN'S Q 464
19-7 STATISTICAL TESTS AND CONFIDENCE INTERVALS FOR THE BIVARIATE NORMAL CORRELATION MODEL 465
19-8 SUMMARY 465
EXERCISES 470

20 INTRODUCTION TO STATISTICAL REGRESSION THEORY 473

20-1 CHARACTERISTICS OF REGRESSION STUDIES 474
20-2 EXAMPLE OF A REGRESSION PROBLEM 476
20-3 ESTIMATION OF THE REGRESSION LINE · METHOD OF LEAST SQUARES 479
20-4 PROPERTIES OF THE LEAST-SQUARE ESTIMATORS 483
20-5 THE VARIANCES OF THE SAMPLING DISTRIBUTION OF $\hat{\beta}$ AND $\hat{Y}_X$ 485
20-6 CONFIDENCE INTERVALS FOR THE UNKNOWN PARAMETERS 488
20-7 TESTS OF HYPOTHESIS FOR THE REGRESSION MODEL 489
20-8 REGRESSION THEORY WHEN THE X VARIABLE IS NOT SPECIFIED IN ADVANCE 490
20-9 EXPLAINED VARIANCE FOR REGRESSION MODELS 493
20-10 REGRESSION AS AN ANALYSIS-OF-VARIANCE MODEL 494

20-11	THE *K*-SAMPLE REGRESSION MODEL (THE ANALYSIS OF COVARIANCE)	497
20-12	TESTS AND CONFIDENCE INTERVALS FOR REGRESSION MODELS	508
20-13	SUMMARY	511
	EXERCISES	516

APPENDIXES **519**

A	LIST OF STATISTICAL TABLES	521
B	LIST OF STATISTICAL TERMS	558
C	ANSWER KEY	564
D	SUPPLEMENTARY READINGS IN RELATED TEXTS	572

INDEX **575**

STATISTICAL METHODS FOR BEHAVIORAL SCIENCE RESEARCH

INTRODUCTION TO MODERN STATISTICS

U.S. SCHOOLS EXPECT RECORD ENROLLMENT

Washington. An estimated record enrollment of about 56 million is in prospect for the nation's public and private schools and colleges this fall, the Office of Education reported Saturday night.

This will be the 22nd consecutive fall enrollment record. The increase is figured at about 2.6 percent, with last year's enrollment estimated to have been 54.5 million.

The office figured that about three of every 10 persons in the United States will be in school at some level this fall—and the cost of operating all schools will push close to $50 billion.

The biggest percentage increase in enrollment is expected in higher education. About 6 million students are expected to enroll in colleges and universities, 9 9 percent more than last fall's 5.5 million.

High school students are expected to increase by 2.3 percent, to 13.3 million from 13 million.

Grade school pupils, kindergarten through grade 8, are estimated to total 36.6 million, up 1 7 percent from 36 million a year ago.

1-1 WHAT IS STATISTICS?

Customarily, first course books on statistics and statistical methodology begin with a definition of the word *statistics*. While this seems like a natural beginning, it is, unfortunately, not an easy one. A major reason for this difficulty is that in recent years the subject matter of statistics has branched out in a number of different directions. As a result, definitions of this term vary, different authors emphasizing one function or meaning in preference to another. Even though there exist variations in meaning among the many accepted definitions of this term, a careful reader of statistics texts would soon note that certain recurring themes or elements appear throughout most of the definitions that have been proposed. While it is not possible, at least at this time, to provide beginning students with a universally accepted definition of statistics, it *is* possible to extract those elements common to the proposed definitions and use them to explain what statistics will mean in the context of this book. As one reads, learns, and gains insight into the subject matter presented, one should eventually come to view statistics as a collection of methods that help an individual or scientist make reasonable decisions from limited amounts of information.

Because this idea or function of statistics is of recent origin, it is not what the layman normally thinks of when the word *statistics* is mentioned. Many people, as well as beginning students, associate statistics with lists of numbers and with esoteric mathematical formulas that permit one to prove almost anything that is desired. In a certain sense these associations are not far from being correct.

For one thing, lists or tables of numbers are compilations of statistics. For example, the tables and charts published by the United States Office of Education on the number of children enrolled in elementary schools during an academic year constitute national statistics on school enrollment. These are, indeed, tables of statistics. They are examples of simple counts or enumeration statistics, the simplest type of statistics collected and reported by data analysts and researchers. The Associated Press release appearing at the beginning of this chapter is an example of the type of statistics with which the layman is familiar.

Another example of the sort of thing that the layman identifies with statistics is the proportion of adults in Los Angeles or San Diego, California that are divorced or widowed. These are also statistics, but they differ from enumeration statistics in that they are mathematically derived from them. In particular, the proportion of adults that are divorced or widowed in these two cities can be found by taking the ratio of the total number of adults who are divorced and widowed to the total number of adults who live in these cities. Of course, the determination of these ratios is not easy. Fortunately, at census years the United States Bureau of the Census collects the necessary information, computes the appropriate ratios, and reports these important social statistics for the nation as a whole.

Another example of the sort of thing people call statistics is the total number of gallons of beer drunk in a particular year. This is an example of a statistic that is found by measuring the beer consumption of the region of immediate interest.

From this one can produce a derived statistic such as the average amount of beer drunk per person in the region. For this measure, it is only necessary to divide the total number of gallons of beer consumed in the year by the number of people who could conceivably drink the beer. In a like manner one could determine the average amount of money spent per person per year for alcoholic beverages in a city like San Francisco, California. If this is then compared to the average amount of money spent per year per person for education, in the same city, the discrepancy in the two statistics might appear to be quite astonishing.

The layman is also correct when he associates statistics with mathematics and mathematical formulas. For the most part, statistics as a formal intellectual discipline is included in the large subbranch of mathematics called probability theory and therefore has its own formulas and equations. However, the layman is significantly wrong in thinking that statistics is esoteric and can be used to prove anything. As the beginning student will learn as he studies this text, statistics is essentially the application of common sense to the solution of simple and complex problems of the everyday world that possess certain regular properties. Furthermore, he will learn that it is not statistics that prove anything but that human beings can deliberately or accidentally make inappropriate use of statistical methods to make false statements. In general, the fault does not lie in the statistical methods that one can use, but rather in the human users of the methods.

To illustrate this point consider the comparison of the average per capita expenditures in San Francisco for alcohol and education. Regardless of what the final statistics are for these two characteristics, there are certain to be citizens who will argue that too much is wasted on alcohol while others will argue that the amount spent on alcohol is only a drop in the bucket. Both groups can start with the same evidence and yet arrive at different conclusions. Depending upon the point of view of the interpreter, statistics can seemingly be used to defend any position, and thereby the layman's belief that statistics can be used to prove anything is true. It should be noted, however, that it is not the statistics that prove either position: the only thing that statistics does is to show how the per capita measures should be obtained and made. In the final analysis, the interpretation and the use of the resulting statistics is in the hands of the observer. Fortunately, statistics has developed certain rules and methods that, once accepted by two individuals, will lead both of them to the same decision and perhaps the same interpretation of what they observe.

1-2 AN ILLUSTRATION OF THE USE OF MODERN STATISTICAL METHODS

As an illustration of the use of statistics, consider a research study in which an event is to be repeatedly observed over a number of consecutive trials made under similar but unique independent conditions. For this study, consider a behavioral scientist who has a preconceived notion that the children in the school district in which he works tend to be lower in abstract reasoning ability than children throughout the nation. His reasons for holding this belief might be that the school district in which he is employed has an overabundance of culturally deprived children. Since it is

common knowledge that the "average" IQ in the national community of children is about 100 IQ points, this behavioral scientist is postulating that in this school district the "average" IQ is below 100.

In order to test his preconceived idea he decides to select a third-grade child from the school district and administer to the selected child the Stanford-Binet intelligence test. The event he plans to observe is the IQ score obtained by the child on testing. Since one observation contains only one bit of information about the IQ level of all third-grade children in the school district, he plans to repeat the experiment with another child, and with another child, etc. Since observations in the school environment cost money and time, he further plans to terminate the experiment after 50 trials and to study the statistics that have been generated. Finally, he intends to make a statement as to what the collected statistics reveal about the IQ level in the school district.

Certainly the individual IQ scores are not of much interest in themselves, except perhaps to the parents of the children involved. However, collectively they are of interest to the behavioral scientist. Although only some of the third-grade children in the school district are to be observed, the behavioral scientist wants to extrapolate from these results a decision about the IQ level in the entire district, thus saving both time and funds. Modern statistical methods enable him to make this jump in knowledge from *some* to *all* within specified limits of error.

Prior to the collection of the IQ measures or statistics, everyone would agree that the behavioral scientist's conjecture is either a true statement or it is a false one—it can never be partially true nor can it be partially false. Only by observing all of the children in the school district could one know for certain whether his conjecture is true or false. However, since only a sample of children is to be selected for the entire school population and since an extension of *some* to *all* is to be made, it follows that a wrong decision might be made on the basis of what is observed. If this were to occur, it would, indeed, be unfortunate.

To illustrate what this means suppose it were true that the conjecture was valid. Prior to selecting the students for the study he would expect most, but not all of them, to have an IQ below 100, and therefore, he and the school superintendent might decide to say that his conjecture was substantiated if the number of children in the sample of 50 with an IQ below 100 exceeded 35. Suppose, in addition, that he observed 42 children with an IQ below 100. Then by direct application of his preselected decision rule he would conclude correctly that the children were lower in abstract reasoning ability. However, it is possible that an unrepresentative selection of children might be obtained even if the conjecture is true, and it might happen that only 12 have IQ scores below 100. In this case his decision rule would lead him to conclude incorrectly that they are not less bright; he would then make a wrong decision in which he denied the truth of a valid conjecture or hypothesis.

On the other hand, suppose the behavioral scientist's guess about the IQ level of the students in the school district had been false. In this case, if 12 children had IQ scores below 100, he would make the correct decision in that he would deny

the statement of his conjecture in favor of the true abstract reasoning ability of the school district's students. However, even though his conjecture is false, it is still possible to have 42 observations below 100. In this case he would make a wrong decision of another kind in that he would affirm the truth of his false conjecture or hypothesis.

The important point to learn from this simple example of decision making with limited information is that the person who engages in this kind of inductive reasoning faces the possibility of committing one of two possible errors: in the one case, it might be a decision that is in agreement with the basic hypothesis but opposed to reality; in the other case, it might be a decision that is opposed to the hypothesis when indeed the conjecture was correct. The possibilities of making these errors are not pleasant to think about, especially if the consequences following the decisions are costly. In the first case, the consequences of declaring that children are just as bright when they are not might mean a financial loss of federal funds to initiate compensatory education programs designed to raise the IQ level of these children, provided, of course, that IQ can be affected by environmental changes. On the other hand, false verification of the conjecture might be just as harmful. This is particularly true if the students are above average intelligence and if money is spent unwisely to attempt to increase the IQ level of children who do not need the extra help. In the first case, the community may suffer an intelligence loss because the native abstract reasoning ability of its future citizens remains undeveloped. In the latter case the consequences of making a wrong decision is to waste federal funds foolishly. If the psychologist of this hypothetical study would study the consequences of these actions and decide which error is more costly, wasted student potential or wasted federal funds, he could then use modern statistics to redesign his study so as to minimize the risks of these errors.

One of the major contributions of modern statistics, and perhaps its major justification for being, is that it provides methods that make the risk of such unfortunate decisions unlikely, and so in this sense statistics is related to a collection of procedures that may be used to help make decisions in the face of uncertainty so as to minimize the harmful effects of the consequences that arise from acting upon the decisions.

1-3 STATISTICS AS A RESEARCH TOOL OF BEHAVIORAL RESEARCH

The procedures that are available to the behavioral scientist to aid in decision making are extremely numerous and the number is continually increasing. In fact there are some statisticians who believe that some of the methods presented in this book are obsolete. If this is indeed the case, and some evidence suggests that it is, then it becomes important that behavioral scientists understand the rationale behind currently used statistical procedures so that transfer to new methods is facilitated. This book is structured with this point of view in the forefront. Since it is an introductory text, only the most frequently employed procedures are presented. As will be noted, this includes methods and procedures for collecting statistics (definitions

as used by the layman), for tabulating statistics, for summarizing statistics, and for interpreting statistics.

Some of the material to be presented is difficult to understand. This is especially true of the material presented in the first half; the second half is much easier, provided that the concepts presented in the first half are understood. It may appear that the concepts in the first half of the book are presented without rhyme or reason; this is not true. Wherever possible, each new idea is built upon those developed earlier and a spiralling presentation of new material is employed. As a consequence, the reader should attempt to master the first half of the book in order to make the second half easier to learn, and, more important, to make it meaningful for behavioral research. To facilitate this learning, the reader is strongly advised to study and re-study old material as new material is being presented. In this way, the necessary connections and associations for a firm understanding of statistics have a higher probability of occurrence. As they occur, the importance of statistics as a research tool for behavioral studies will become evident. While statistics is not required for all research problems, its utility in studying many significant questions of behavioral research is well documented by the hundreds of research publications that have used statistics successfully. Without doubt, its continued use as a research tool of behavioral scientists is assured.

1-4 WHAT STATISTICS IS ABOUT

The fundamental concept of statistical theory is that of a population or universe of objects, either real or hypothetical. The primary function of statistical theory, as it will be presented here, is to formulate and develop methods that will enable scientists to make valid and reliable statements about some measurable characteristic possessed by the objects or elements of a population on the basis of a sample selected from a population or universe under study. These statements usually involve estimating the numerical values of the properties possessed by the variables being studied and making inferences about the relationships that the variables under study have to other variables or characteristics also possessed by the objects of the universe. A clarification of these statements is presented in the following examples.

Consider the population of students enrolled at the University of California in the fall quarter of a specified year. Interesting characteristics possessed by the members of this population are their sex, their age at entering college, their high school grade-point average, their attitudes toward student demonstrations, the religious preference and political affiliation of the students and their parents, and many others. By selecting a small, statistically designed sample of Cal students, one could estimate certain summary measures on a small sample of the students and then make statements about differences in attitudes toward student demonstrations according to the sex, age, or grade-point average of the students, or according to the religious or political affiliations of their parents. Furthermore, one could determine what relationship the age at entering college has to the attitudes possessed by students toward demonstrations or to any other variable of interest.

Another interesting variable associated with the male members of the population of University of California students is their height on entrance and their high school basketball performance as measured by the number of points they scored in high school varsity games. The Cal basketball coach could select a sample of students who had played on high school basketball teams and observe in the sample the relationship that these variables have to one another; he could then make an inference as to the total population of incoming male students who had played on high school basketball teams. With luck, diligence, and good selection practices he might produce a first-rate basketball team, which could go on to gain national fame.

Consider the population of unwed mothers living in a lower-class neighborhood of a specific large Northern urban community. Interesting characteristics possessed by the members of this population are race, number of completed years of education, age at first sexual union, age at delivery of first child, yearly income, usual occupation, early childhood experiences, and the decision to raise their own illegitimate children or the decision to put them up for adoption for someone else to raise. To study the variables associated with the decision to raise their own offspring, a social welfare researcher could select a small sample of unwed mothers, obtain information on these variables through interviews or records, and then study or determine the relationships that the decision has to early home experiences, number of completed years of education, race, and age at first sexual union. After looking at such relationships and variables he could make general statements about the population of unwed mothers living in lower-class neighborhoods. Hopefully, he could then recommend changes or modify existing relief and aid programs to assist and reduce the social problems created by unwed mothers.

Consider the population of third-grade children that is being taught arithmetic in the present academic year. Interesting characteristics possessed by the members of this population are their age at the beginning of the school year, their race, their attitude toward their teacher, their socioeconomic background, and their success in learning arithmetic as evidenced by scores on standardized achievement tests. By selecting a small sample of third-grade students an educator might wish to determine whether the amount of success attained in learning arithmetic is a function of the age at which a student enters the third grade, the race of the student, his attitude toward his teacher, or his socioeconomic background. Finally, by studying the interrelationships that these characteristics have toward one another in the sample, the investigator might be able to modify existing teaching methods to help poor achievers in arithmetic.

Consider the population of sophomore psychology students traditionally victimized into participating in psychological experiments. Frequently in such studies the population of sophomores is partitioned into what are called a control population and an experimental population. As a typical experiment, the students in the sample selected from the control population might be told in a laboratory setting that they are expected to learn a list of 10 pairs of nonsense words in as few trials as possible. Of course, they are not told that almost all students learn the list

in 15 trials. The students selected for the experimental condition may be told the same thing but they are also told that if they don't learn the list in 20 trials they will be given a mild electric shock after each trial until they can give the complete list from memory. So as to further increase the anxiety level of these students, the testing laboratory may be stocked with meters, wires, and other electrical paraphernalia to which the subjects of the experimental population are wired. Interesting characteristics possessed by members of these two populations are their sex, their normal state of anxiety as measured by a verbal test of anxiety given prior to entering the laboratory, the number of trials it takes to learn the list, and their GSR (galvanic skin response) measure. By means of a small sample, a psychologist can determine whether the effects of anxiety differ between males and females, whether subjects with generally low levels of internal anxiety respond differently to the stimuli than do subjects with high levels of internal anxiety, and finally whether there is an interaction between sex and internal anxiety as measured by the GSR and the number of trials required to learn the lists under the two conditions of the experiment. Depending upon what is learned, the psychologist's knowledge of the effects anxiety has upon learning is extended or increased.

Consider the population of adults living in a community in which the school board has decided to make administrative changes in the schools to achieve school integration. Interesting characteristics possessed by the members of this population are their race, the number of years lived in the community, their involvement with the schools to be integrated, their attitudes toward integration, and their socioeconomic background. A sociologist might decide to sample this community and measure these characteristics on the sample members. He can study the sample responses to determine what connections race, number of years lived in the community, involvement in the community, and socioeconomic background have with attitudes toward integrating the schools. Not only will he learn something about the effects that these variables have upon one another but he might also be able to make meaningful suggestions to administrators in other school districts that are contemplating similar changes.

Other examples could be listed *ad nauseam*. However, there is no need to make such an extensive listing, since these six examples illustrate the basic elements of most statistical studies. There is always a population that is sampled. Within the sample certain kinds of measurements are made upon the variables or characteristics that are of interest to the researcher. The variables are studied and the interrelationships existing between them are observed. From this, an inference or set of inferences about the entire population is made.

It should be noted that the populations studied need not be real populations in the sense that a census or count could be taken of all elements in the population. The population of University of California students is a real population. The registrar at the university could prepare a list of all students registered in the fall quarter of the year specified by the study.

The population consisting of the male members of the university who had played on high school basketball teams is also real. Generally, it would be true that the basketball coach who makes such a study would not be interested exclusively in the population of California male students. In the typical case he would be looking for a relationship that holds for all male students and over many years of time. For him the population of interest is hypothetical and somewhat abstract.

This is even more true for the population of unwed mothers. The sociologist of this study is obviously interested in determining general statements of truth for unwed mothers, not just for the unwed mothers of the city where he is doing his experiment. His interest is in all unwed mothers past, present, and future who live in large urban areas. His population is hypothetical and is quite abstract.

The population of third-grade children is similar. While the school superintendent could provide a list of all third-grade children enrolled during the present academic year, the educator conducting the study is interested in making statements that hold for third-grade students in general. Such a population is an abstraction.

The same is true of the population of college sophomores. At any one time there is a group of sophomores who enroll to take introductory psychology and thereby make themselves possible guinea pigs for psychological experimentation. Again, the psychologist is not interested in making inferences about the 600 students who may be enrolled in his class in psychology. Furthermore, his interest in sophomores is not the strongest. Instead, he wishes to make statements about the way "man" reacts to anxiety under the situations that a psychologist creates in the laboratory. In this sense, he is concerned with an abstract population of men past, present, and future. Hopefully, the fruits of his research will lead to a better understanding of anxiety and how it affects man's behavior.

Finally, the population of adults is a real population that could never be enumerated except through a population census such as the ones conducted in decennial years by the United States government.

1-5 KINDS OF STATISTICAL VARIABLES

As seen from the six examples of Section 1-4, the kinds of variables in which a scientist might show interest cover a wide spectrum of choice. Fortunately they can be classified into rather broad rubrics, which are useful when selecting appropriate statistical methods. To help differentiate among these broad forms consider some of the variables named in the six examples. In doing this it is convenient to adopt the statisticians' custom of denoting variables by the capitals of the letters near the end of the alphabet. Since the classification scheme will involve three broad rubrics, let the letters that will be used be X, Y, and Z, with appropriate subscripts to distinguish between different variables contained in the same rubric. Before the classification scheme is discussed, try to see what the X variables, the Y variables, and the Z variables have in common. The variables are listed in the order in which they appear in the examples.

X: Qualitative Variables
- X_1: Sex of student
- X_2: Attitudes toward student demonstrations
- X_3: Religious preference
- X_4: Political party of student and parents
- X_5: Race of mother
- X_6: Usual occupation
- X_7: Early childhood experiences in parents' home
- X_8: Behavior toward adoption
- X_9: Race of the student
- X_{10}: Attitudes toward teachers
- X_{11}: Socioeconomic background
- X_{12}: Involvement with the schools
- X_{13}: Attitudes toward integration

Y: Quantitative Discrete Variables
- Y_1: High school grade-point average
- Y_2: Number of basketball points earned in varsity games
- Y_3: Number of completed years of education
- Y_4: Yearly income
- Y_5: Score on standardized achievement test
- Y_6: Score on anxiety test
- Y_7: Number of trials to learn list

Z: Quantitative Continuous Variables
- Z_1: Age at entering college
- Z_2: Height on entering college
- Z_3: Age at first sexual union
- Z_4: Age at delivery of first child
- Z_5: Age at beginning of school year
- Z_6: GSR (galvanic skin response)
- Z_7: Number of years lived in community

The obvious feature associated with the X variables, such as sex of the student, attitudes toward student demonstrations, religious preference, political party of the student and parents, is that the variable may be partitioned into nonoverlapping and exhaustive classes. For example, sex can be divided into two nonoverlapping and exhaustive classes: {male, female}. Race can be partitioned into {Caucasian, Negro, Oriental, other}. Religious preference is frequently partitioned into {Protestant, Catholic, Jewish, other}. Variables which can be partitioned into a set of nonoverlapping and completely exhaustive classes are said to be attribute or *qualitative* variables. They specify a characteristic that is defined by the set of classes that comprise it.

Sometimes the classes of a qualitative variable carry a ranking or ordering pattern.

For example, college grades are usually reported as $\{A, B, C, D, \text{and } F\}$ where it is understood that an A is a higher ranking than a B and that an F is the lowest ranking that a student could be assigned. Attitudes toward student demonstrations are usually assigned to the classes of an order ranking frequently of this form: {highly favorable, moderately favorable, neutral, moderately unfavorable, highly unfavorable}. Both of these examples are ordered qualitative variables. Variables such as race, sex, eye color, occupation status, etc., are qualitative variables that lack an ordering property.

Finally, it should be noted that qualitative variables are probably the most frequently encountered variables or characteristics studied by behavioral scientists. They are also the most difficult to analyze from a statistical point of view. Some of the elementary statistical procedures that may be used in their analysis are presented in Chapters 16 and 17.

The sets of Y and Z variables do not contain the same classification property as that possessed by the X variables of the six examples. These variables are clearly numerical in nature and are called *quantitative variables*.

The Y variables, such as the number of completed years of education, yearly income, score on a standardized achievement test, score on an anxiety test, and the number of trials to learn a list, have a counting property about them. One can count the number of completed years of education had by an individual. The number of pennies that a person earns in a year can be counted. The number of correct items on a standardized achievement test can be counted. The number of items reflecting anxiety on a test can be counted. The number of trials it takes to learn a list of nonsense words can also be counted. In every case, one could list in an increasing order the possible values that the variable or characteristic can assume. For example, the possible values for the number of completed years of education is given by {0 years, 1 year, 2 years, 3 years,...}. Yearly income can include any of the numbers {\$.00, \$.01, \$.02,..., \$6,532.75, \$6,532.76,...}. In this last example, it should be noted that no one can ever earn an income that is not on the list. For example, no one can have an income of \$6,532.755, since \$.01 is the basic unit of measurement of American currency. All of the Y variables have this property; in-between values are not possible. Variables that possess this property are called *discrete variables*. They can assume only certain distinct numerical values. In theory, the values they can assume can always be listed.

Discrete variables can assume a finite set of values or they may assume an extremely large set of values for which it is impossible to specify the last or largest value. For convenience the latter type of discrete variable is said to assume an infinite set of values. All six of the discrete variables presented in this section are examples of variables with a finite range of values. Finally, it should be noted that discrete variables account for most of the quantitative variables encountered in the behavioral sciences. Reasons for this will be presented later.

As an added comment, it should be noted that a discrete variable with an infinite set of possible values is a pure mathematical abstraction. Such a variable could not

exist in the real world of experience. The number of stars in the universe is finite; the number of grains of sand along the California coast is finite; the number of original ideas that a person might generate in a lifetime is finite; the number of words that a person utters in a lifetime is also finite. However, in each of these cases the maximum number available for each variable is extremely large and beyond man's comprehension. For that reason, it is convenient to say that they are infinite in extent.

The Z variables, such as age at entering college, height, GSR, etc., are called *continuous variables*. Like discrete variables they are not characterized by class partitioning, but unlike discrete variables they are not determined by counting. Instead, they are determined by measuring. Furthermore, they can assume all possible values in a finite continuous range of values and for this reason they are called continuous variables. For example, a freshman who enters college may have been 18 at his last birthday, but at the time the school year begins he may really be 18 years, 4 months, 12 days, 15 hours, 8 minutes, 27 seconds old. Heights are traditionally measured by comparing a standing person to a set of marks on a yardstick. Between any two marks on the yardstick there is always a place for another mark. So in this sense, height can assume an infinite set of numbers with no empty spaces between numbers. The GSR can also take on an infinite set of numerical values in a continuous range of numbers, the particular value being determined by the amount of electric current that can pass across the surface of the skin at the time of testing.

Many of the variables encountered in the physical and biological sciences are of a continuous nature. Examples are atmospheric pressure, air temperature, lung volume, body weight, wood density, age at death, etc. Even though a variable is continuous, man's measuring instruments are by necessity calibrated in discrete steps or marks, which hopefully are equally spaced. For example, temperature is measured by man in 1° steps even though temperature itself changes in a continuous fashion. This discrete property of measurement is essentially true for all the variables that man measures.

Some continuous variables are encountered in the behavioral sciences but to a lesser extent. Some psychologists believe that intelligence represents a continuous variable, which unfortunately cannot be measured by the instruments available to man at this time. Instead, psychologists obtain a score on an IQ test, which hopefully has some bearing on intelligence. In any case, the IQ score is on a discrete scale of values. Furthermore, some psychologists believe that IQ scores do not increase on a scale for which the steps are of equal value. For example, the difference in abstract reasoning ability between IQ scores of 125 and 130 is not the same as the difference between IQ scores of 105 and 110, even though the numerical magnitudes of the test-score differences are the same in both cases.

Since the statistics of continuous variables is simpler than that of discrete variables, it is convenient to treat discrete variables as continuous variables. Under certain conditions this adjustment can be justified. Whenever one can make the adjustment

or correction for continuity, as it is called by statisticians, one is assuming that a discrete variable with the numerical value Y is contained somewhere in the interval of $Y \pm \frac{1}{2}U$, where $U =$ the smallest measurable unit. Thus, if a thermometer calibrated in 1° steps shows that the temperature is equal to 20°F, one should not hesitate to state that the true temperature is in the range $19.5° < T < 20.5°$, and that the exact temperature is still unknown. In like manner, when dealing with test scores it is customary to associate a test score of 23 with the continuous interval $22.5 < S < 23.5$. The reasons for this conversion will be made clearer later.

To summarize, statisticians characterize variables as qualitative or quantitative with the quantitative being further partitioned into discrete and continuous variables. Knowledge of this classification is useful since it simplifies the selection of appropriate statistical procedures. The first thing to note when making a decision as to what statistical procedure should be used in studying a particular problem is the statistical nature of the variable. Once this is determined, the appropriate procedure is essentially known.

1-6 SUMMARY

The original function of statistics was to serve kings, rulers, and magistrates. The earliest use of statistics was to count the number of citizens who were available for service in the army or who could be taxed to support the state. According to biblical accounts, Mary and Joseph were returning to Bethlehem in Judea to be counted for the Roman census and tax year ordered by Caesar Augustus. In 1787, the founding fathers of the United States decided that a census of the population should be taken every ten years so that the statistics generated from the census could be used to determine the number of Congressmen each state could send to the House of Representatives in Washington. This practice is still in operation today. As science and technology began to emerge and grow, statistics began to change and develop so that today one of its major functions is to help individuals make decisions on the basis of limited information. Generally, the information is obtained from small samples selected from a much larger set of elements called a population. Observations are made on the objects in the sample, relationships in the sample are studied and tested, and then inferences are made on the population from which the sample came.

In general, the observations are taken on characteristics or variables possessed by the elements of the sample. These variables are of three kinds: qualitative, discrete, or continuous.

1. Qualitative variables are essentially classification variables. A qualitative variable can assume any one of nonoverlapping and completely exhaustive classes. The classes can be ordered or else they can be quite unrelated to one another. An example of an ordered qualitative variable is given by size of urban centers: {0 to 2,499; 2,500 to 9,999; 10,000 to 49,999; 50,000 to 99,999; 100,000 to 499,999; 500,000 to 999,999; 1,000,000 and over}. An example of an unordered qualitative variable is given by hair color: {blonde, brunette, red, other}.

2. Discrete variables are essentially counting variables. A discrete variable can assume any one of a finite or countable set of unique values. An example of a discrete variable is the number of times an automobile driver is given a traffic ticket over a specified period of time for violating local parking laws. This variable can assume values $\{0, 1, 2, \ldots, K\}$, where K is the maximum number of tickets given to any one member of the population of automobile drivers.
3. Continuous variables are essentially measured variables. A continuous variable can assume all possible numerical values in a finite or infinite range of continuous values. An example of a continuous variable is the amount of time patients spend on an operating table for the removal of cancerous tissue. This variable can assume values $\{0 < t \leqslant T\}$, where T is the maximum amount of time that any one person of the population of individuals who undergo such operations spends on the operating table.

Most of the variables encountered in behavioral research are qualitative or discrete. Continuous variables are less commonly encountered because present-day measurement techniques do not permit perfect measurement. Since most of the classical statistical procedures assume that variables are continuous in nature, it is generally necessary to make corrections for continuity. One of the ways to make this correction is to assume that if a variable has a measure of Y units, one can behave as though the exact value is unknown but is in the range of $Y \pm \frac{1}{2}U$, where U is the unit of measure. This procedure will suffice for most of the methods to be presented in later chapters.

EXERCISES

1-1. Distinguish between enumeration statistics and derived statistics. Give examples of each.

1-2. It is argued that statistics lie. Is this a valid statement? Why?

1-3. A statement about the external world is true or it is false, it is never partially true. Explain why this is so.

1-4. In American courts of law, a person suspected of a crime is assumed innocent until proven guilty. What are the possible errors that may result from a court trial under the American system of passing justice? Which of these errors is considered to be the more serious? How does American society protect itself against making the more serious of the two kinds of errors?

1-5. Distinguish among continuous, discrete, and qualitative variables by means of examples.

1-6. Many of the universes of behavioral research are hypothetical in nature. Give some examples of real and hypothetical universes that may be encountered in behavioral research. On these universes define a continuous, a discrete, and a qualitative variable.

1-7. Classify the following variables as to whether they are qualitative, discrete, or continuous. Give reasons for your choice.

(*a*) Age at entering college.
(*b*) Years of completed education.
(*c*) Graduate record examination scores.
(*d*) IQ as measured by the Stanford-Binet intelligence test.
(*e*) Eye color.
(*f*) Number of wrong turns that a rat makes while running a complex maze.
(*g*) Responses to an attitude questionnaire where the choices are: strongly disagree, moderately disagree, moderately agree, strongly agree.
(*h*) Time it takes to complete a task in a learning experiment.
(*i*) Score a student obtains on a reading test.
(*j*) Weight of a guinea pig before the start of an experiment in which vitamin B_1 is to be deleted.
(*k*) Number of arrests for a misdemeanor prior to an arrest for the commission of a felony.
(*l*) The percent of correct answers on a 15-item test.
(*m*) Letter grades *A*, *B*, *C*, *D*, and *F* assigned to a term paper on the importance of statistics in behavioral research.
(*n*) Distances reached by standing broad-jumpers.
(*o*) Atmospheric pressure at noon during the month of June in Chicago, Illinois.
(*p*) Amount of money spent on food per week by a family of four people.
(*q*) The distance traveled by a guinea pig on a rotary treadmill in a 24-hour day.
(*r*) The proportion of voters who support a particular candidate in a senatorial election.
(*s*) The proportion of proteins in the daily food intake of an immigrant to the United States.
(*t*) Self-esteem as measured by the total score obtained on an inventory of 10 items, each extending over a five-point rating scale.

1-8. Explain and illustrate by means of an example what is meant by an ordered qualitative variable. Contrast it with a quantitative and a continuous variable. Explain how it differs from a discrete variable.

1-9. Even though a variable such as height on twenty-first birthday is continuous, man's measuring procedures create a variable for analysis that is discrete. Explain.

1-10. Of what value is a discipline such as statistics, which is concerned with arguing from *some* to *all*? Defend your position by an example.

2 INTRODUCTION TO SET THEORY IN BEHAVIORAL RESEARCH

.. in modern times speculations about dreams have been long on theory and short on observation. What do people dream about? What is the content and character of their dreams? As far as I know, no one has made an extensive and systematic study of these questions....

The writer has undertaken to make ... a survey, for the purpose of obtaining some empirical facts as a foundation for theorizing. He has collected more than 10,000 dreams thus far, not from mental patients but from essentially normal people. They were asked to record their recollection of each dream on a printed form which included questions requesting certain specific information, such as the setting of the dream, the age and sex of the characters appearing in it, the dreamer's emotions and whether the dream was in color....

We classified the dream material so that it could be studied statistically. From the many possible methods of classification we chose, as a beginning, a simple breakdown into five fundamental categories 1) the dream setting, 2) its cast of characters, 3) its plot, in terms of actions and interactions, 4) the dreamer's emotions, and 5) color.

2-1 POPULATIONS

As was suggested in Chapter 1, the primary function of modern statistics is to help researchers make inferences from a small set of observed data to a larger set of unobserved data. In terms of the language employed by modern statisticians, the small set of observable data is called the *sample* while the larger set of unobserved data is called the *population* or *universe* of investigation. Populations of persons, places, or things should not be strange or unusual to the readers of this book, since these are generally among the first concepts that are learned as infants. In essence, populations comprise the classes or nouns of daily existence. For example, all living persons are intimately acquainted with the population of human beings living on the surface of the earth at the present time since they, themselves, are members of it. Other populations with which most individuals have at least a passing knowledge are the population of children attending school in the local school district, the population of women who undergo illegal abortions in a specified community in a given year, or the population of thoughts that an individual has on any day of his life. Some of these populations can be touched and felt and measured directly; some must be observed and measured indirectly from great distances; and some will remain inscrutable and immeasurable for all time.

While these populations might be of interest themselves, intellectual curiosity generally focuses on certain characteristics or variables possessed by the members of these populations. For example, in the population of human beings living in the United States, a sociologist studying human behavior might find that leisure activities and the amount of money spent per year on entertainment are interesting attributes for study and theory building. For the other populations, interesting variables of study and scientific investigation might be the IQ scores obtained by high achieving children given the Wechsler intelligence test, the racial composition of women who commit abortions, or the proportion of thoughts a college sophomore might have in a particular day that are indicative of a creative mind.

For the statistical study of dreams discussed in the introductory passage of this chapter, classification into (1) dream setting, (2) cast of characters, (3) plot, (4) dreamer's emotions, and (5) color was required before analysis could begin. Dream settings such as open or closed space, presence or lack of water, night or day, stable or unstable surroundings, and other kinds of settings had to be defined and agreed upon as having relevance to the study of dreams before the analysis could proceed. The setting up of such domains of study is generally the first step to data analysis. Once these population characteristics are defined, statistical tabulations and analysis proceed directly.

As is well known, characteristics vary from member to member in each of the particular populations. For this reason statisticians call population characteristics *variables*; furthermore, the entire collection of these numbers or classes are referred to as populations of measures or classes Thus, while it is appropriate to refer to the students in a particular school district as a population, it is also correct to refer to their IQ scores as a population of IQ measures. In a similar fashion, it is correct to

speak of the population of dollars spent per annum by citizens of the United States on entertainment, the population of ages of women who submit to abortions, or the population of the percentage of thoughts in a particular day that are indicative of a creative mind. As these examples suggest, a population or universe consists of all possible numerical values or exhaustive classes that a certain variable can assume over the collection of objects under study.

To further illustrate this notion, consider the population of women who undergo illegal abortions. Some are Caucasian, some are Negro, and some are Oriental. If these three classes were to exhaust the races of the earth, one could essentially assign all women who undergo abortions to these three classes. The women who constitute the members of these classes are called subpopulations or subsets of the universe of women who commit abortions.

As another example, consider a school district with 29,857 students who are given the Wechsler intelligence test. All the children with an IQ of 113 constitute a subpopulation or subset of the entire collection of 29,857 students. The same is true of the students who have an IQ of 125, 87, or any other discrete integral score.

As these examples show, any universe can be partitioned into subsets of elements. This can be accomplished in many ways. For example, not only can school children be partitioned by IQ, but they can also be partitioned by age, sex, race, grade, religion, school, neighborhood, occupation of father, scores on particular achievement tests, etc. This notion of a universe and its related subsets is a basic building unit of modern statistics. Because of this, a useful and yet simple development of the basic ideas of mathematical set theory will be presented. The concepts to be introduced are very simple and are in the repertoire of most everyone. Surprisingly, many individuals find this material difficult on first exposure. Perhaps the major reason for this difficulty is that while the elementary principles of set algebra are really very simple, most people look for more in the theory than is essentially and actually there. This frequently leads to unnecessary confusion and anxiety. While it is true that advanced topics in set theory can be difficult, the material presented here is not of such complex nature. The discussion on sets is restricted to those parts of the theory that are necessary for the understanding of probability and the simple method of statistical inference developed in this book.

As might be expected, there are many ways to argue from *some* to *all*. The method developed and illustrated in this book is only one possible method for doing this effectively. Perhaps the major justification for its usage over other methods is that it permits researchers in various parts of the world to come to the same decision, provided that their experimental or study conditions are similar. Other justifications for the use of these methods are to be found in their simple logical base, the ease with which they may be executed, and their workability and meaningful results. However, it must not be assumed that the methods developed and described here are the ultimate; with the important research and expansion of methodology being conducted by present-day mathematical statisticians, one should not be surprised if in 20 years the methods described in this book are considered old-fashioned. The

theory of statistics is evolving rapidly and it is for this reason that an understanding of the rationale behind modern statistical methods is needed by researchers in the behavioral sciences. Newly introduced methods will be easier to comprehend and will be used properly with greater frequency if the researcher has a good understanding of present-day techniques.

Some individuals may find the algebra of sets challenging and interesting while others will surely think that it has been inadequately treated here. Both types of students may extend their knowledge by referring to any one of the many good texts available that present more of the advanced and esoteric portions of the theory of sets and its algebra.

2-2 INTRODUCTION TO SETS

To introduce the notion of a set it is convenient to start with its everyday common-usage meaning. Thus, for a beginning, let a *set* be defined as a simple aggregate or collection of distinct objects of any sort. Let the individual objects of the set be called the *elements* of the set. The entire collection of distinct objects will be said to constitute the set.

For example, the set of English letters that constitute the word "Mississippi" is given by the four distinct letters that make up the word. S: $\{m,i,s,p\}$. The order in which the elements are listed is arbitrary. Consequently, when listing the elements of a set, it is customary to ignore the order in which the elements are listed; therefore, S: $\{i,m,s,p\}$ or S: $\{p,m,i,s\}$ are equivalent statements describing the same set. As this example suggests, a set is denoted by a capital English letter such as S, A, B, followed by a colon, followed by a brace and a listing of the elements, and terminated by a closing brace.

The 29,857 students of the previously discussed school district constitute a set of students that could be listed in alphabetical order on a name roster as S: {Adina Abbott, Ben Adam,..., Zerbinetta Zwickert}. Perhaps of greater interest would be the population of Wechsler IQ measures associated with these students. If, for example, the IQ of Adina Abbott were 112, that of Ben Adam were 119,..., and that of Zerbinetta Zwickert were 98, then the set of IQ scores could be summarized by the set S: $\{112, 119, \ldots, 98\}$. In this particular set the numbers 112, 119, and even 98 would be repeated many times because of the large number of students enrolled in the school district. Rather than report the same IQ score many times, it is customary to list only the distinct IQ measures. This can be done by starting with the lowest score and ending with the highest score. Thus, if the lowest score were 72 and if the highest score were 179, then the set of IQ scores could be indicated as follows: S: $\{72, 73, 74, \ldots, 178, 179\}$.

Other sets are given by:

1. The set of oceans of the earth, S: {Antarctic, Arctic, Atlantic, Indian, Pacific}.
2. The set of the number of spots appearing upon the upper face of a balanced die that is shaken and thrown from the hand, S: $\{1,2,3,4,5,6\}$.

3. The set of totals that appear when two balanced dice are tossed as in the game of craps played in the gambling casinos of Nevada, S: {2,3,4,5,6,7,8,9,10,11,12}.
4. The set of integers, S: {1,2,3,4,...}.

Since the fourth example does not have a terminal number, it is customary to indicate this by placing three dots after the first few elements of the set under discussion. Such sets are very common in mathematics but probably nonexistent in the real world encountered by behavioral scientists. However, this does not mean that behavioral scientists should ignore them. Frequently, they can be used to simplify the analysis and understanding of complex behavioral problems. Many examples of infinite sets will be encountered in the remaining pages of this book.

Finally, one should note that not all sets can have their members listed on a roster. Sometimes they must be described verbally or else a range of possible values must be specified. For example, the set S: {thoughts that a person has on a particular day that are indicative of a creative mind} must be stated verbally. However, the set of weights attained by humans at their forty-fifth birthday must be denoted by $S: \{80 < X < 450\}$. The inequality specified in the brackets is a shorthand symbolism for the sentence: X is the set of weights that range between 80 pounds and 450 pounds but does not include weights of exactly 80 or 450 pounds. If one wanted to include the endpoints, then the set defined by the inequality would be denoted by $S: \{80 \leqslant X \leqslant 450\}$. If one wanted to include only the lower end point, then the set would be denoted by $S: \{80 \leqslant X < 450\}$. Other examples of verbally stated sets are: the set of individuals who support a particular candidate in an election, S: {supporters of the candidate}; the set of students who fail a language arts test, S: {failers of the language arts test}; the set of women who commit abortions, S: {women who abort}; the set of tourists who visit Spain in a given month, S: {visitors to Spain}; the set of responses to an attitude question, S: {strongly agree, moderately agree, moderately disagree, strongly disagree}; the set of musical concerts given at Lincoln Center in New York City last season, S: {concerts in Lincoln Center last season}.

As another example, consider the set of totals that can appear when a pair of balanced dice is tossed simultaneously. Clearly, a total of 5 is an element of this set since it is possible to toss a 1 and a 4, a 2 and a 3, a 3 and a 2, and a 4 and a 1. However, it is physically impossible to toss a total of 23 since the possible maximum total is equal to 12. This would occur if a 6 and a 6 were to appear upon tossing. Thus, 23 is not a member of the set of totals that appears when two balanced six-sided dice are tossed simultaneously.

This exclusion of possible set members illustrates one other property that a set must satisfy, and this property is that whenever an object or element is being examined it should be possible to determine unequivocally whether or not the element is or is not an element of the set under discussion. When this can be done, a set is said to be well defined. Thus, a *set* is a well-defined collection of distinct objects or elements of any sort. If A represents a set of objects, and $a_1, a_2, a_3, \ldots, a_i$ represent the individual elements, then the set denoted in roster form is as follows: $A: \{a_1, a_2, a_3, \ldots\}$. If

a_i is an element of A, one writes $a_i \in A$. If a_j is not an element of A, one writes $a_j \notin A$. In the tossing of two balanced dice, if $a_i = 5$, then $5 \in S$, but if $a_j = 23$, then $23 \notin S$.

Finally, it should be noted that a set can be finite or infinite in extent. The set of 29,857 students constitutes a finite set; their IQs also constitute a finite set, as does the set of schools that these students attend. Furthermore, the set of totals that appears when two balanced dice are tossed simultaneously or even the set of oceans of the earth constitute sets with a finite number of elements. However, the set of integers is not finite. It has an infinite number of elements, which can be counted or enumerated; this is why a last element cannot be indicated. The set of weights attained by humans at their forty-fifth birthday is also infinite in extent, but this set of values cannot be counted. It is said to constitute an uncountable set of elements. As these examples suggest, discrete variables will be associated with finite sets or with infinite sets with a countable set of elements, while continuous variables will be associated with uncountable infinite sets. It should be noted that sets are simply the universes or populations of study in behavioral research.

In any behavioral study identical research questions are not directed toward all of the elements of a universe. In a study of college choice on the part of high school graduating seniors, a careful researcher would separate boys from girls and study the two groups of students individually. Under this model, the boys constitute a set and the girls constitute a set. These sets are said to be subsets of the set of graduating high school seniors.

As another example of a set and a subset note that in the set of numbers that appear when a balanced die is tossed, three of the numbers, A: {2,4,6}, are even. These numbers constitute the set A: {even numbers}. This set is a subset of the universe of possible outcomes, S: {1,2,3,4,5,6}. As both of these examples suggest, every set contains subsets. Since researchers usually focus their attention on subsets of universes for investigation, it is useful to consider and elaborate on the idea of a set and its subsets. These notions will then be used to define domains of study in scientific investigations. Finally, they will be used as the building blocks in the construction of a theory of inferences from minimal data.

Most of the examples presented in the following paragraphs involve finite sets. This has been done only to simplify the presentation. While the definitions are stated in terms of finite sets, they apply to countable or uncountable sets. Later in the book, infinite sets will be discussed, but to illustrate the concepts of set algebra, most of the discussion will be restricted to discrete finite sets.

2-3 THE EMPTY SET

A set that has no elements is called an *empty* or *null set* and is denoted by $\emptyset$. The set $\emptyset$ has no members. It is a set with no elements. While it may seem surprising, empty sets are extremely common in the everyday world. Some examples of empty sets are: $\emptyset_1$: {women presidents of the United States}, $\emptyset_2$: {males who are mothers}, $\emptyset_3$: {totals that exceed 12 when two balanced six-sided dice are tossed simultaneously},

and $\emptyset_4$: {human beings who have set foot on Jupiter}. All of these sets are without members and are therefore empty sets, at least at the present time.

2-4 EQUAL SETS AND SUBSETS

The set A: {totals that exceed 10 when two dice are tossed simultaneously} and the set B: {the two largest totals that appear when two dice are tossed simultaneously} contain the identical elements 11 and 12. Even though the sets A and B are defined in different ways they are nonetheless the same set, or equal sets.

As this example suggests, two sets $A: \{a_1, a_2, \ldots\}$ and $B: \{b_1, b_2, \ldots\}$ are said to be *equal sets* if, and only if, they have exactly the same elements. If two sets have this property, one writes $A = B$. The two sets of numbers $A: \{3,9,15,17\}$ and $B: \{3,15,17,9\}$ are equal. The elements of these sets appear in different orders, but every element in one set is also an element of the other set. Note that if two finite sets do not have the same number of elements, they cannot be equal. Equal sets always contain the same number of elements, but it need not be that sets with the same number of elements are equal. For example, $A: \{2,4,6\}$ and $B: \{1,3,5\}$ are not equal sets even though both contain three distinct elements.

The set A: {assistant professors of the University of California} and the set B: {assistant professors of the University of California who were born in California} are not equal sets. Yet the set B is contained completely within the set A. This means that all elements of B are also elements of A but not all elements of A are elements of B. When this happens, it is said that B is a *proper subset* of A and is denoted by $B \subset A$. If there is a possibility that $B = A$, one writes $B \subseteq A$. In this book $B \subset A$ is used to represent the statement "All elements in B are in A." In this sense it is also possible for B to be identical to A.

The set A: {college sophomores} is a subset of S: {the population of college students}. The set A: {John Adams and John Quincy Adams} is a subset of S: {presidents of the United States}. The set A: {Negro} is a subset of S: {races of earth}. The set A: {tails} is a subset of S: {outcomes when a coin is tossed once} = S: {tails, heads}.

2-5 COMPLEMENTARY SETS

The set S: {integers} contains as a subset A: {even integers}. Those integers that are not elements of A, the odd integers, are referred to as the complementary set of A.

Thus, if S represents a universe or population and if A is a subset of S, then all the elements of S that are not in A are called the *complementary set* of A and this set is denoted by $\bar{A}$. The complement of A is a subset of S, that is, $\bar{A} \subset S$.

With reference to the appropriate universe, the complement of {male} is {female}, the complement of {success} is {failure}; the complement of {yes} is {no}; the complement of {lived} is {died}; the complement of {Caucasian} is {non-Caucasian}; and the complement of {rejection} is {nonrejection}, rather than {acceptance}.

2-6 THE INTERSECTION OF TWO SETS

Consider a universe S: {students enrolled at the University of California at the last fall quarter}. The set M: {males} constitutes a subset of that population as does the set F: {freshmen}. Some of the members of the universe are members of both sets. That is, some of the males are freshmen and some of the freshmen are males. This common set I: {male freshmen} is said to constitute the intersection of the set M: {males} with the set F: {freshmen}. This common set of students constitutes a subset of the universe, a subset of the subset of males and a subset of the subset of freshmen, that is, $I \subset S$, $I \subset M$, and $I \subset F$.

More formally, let S represent a universe or population. Let A be a subset of S and let B be a subset of S, that is, $A \subset S$ and $B \subset S$. The elements that are in A and simultaneously in B are said to constitute the *intersection* of the sets A and B. This common set is denoted by $I = A \cap B$.

Thus, if S consists of the following set of finite numbers S: {5,6,7,8,9,10,11} and if A is given by A: {5,6,9,10} with B given by B: {6,9,11}, then the intersection of A and B is given by $A \cap B$: {6,9}. As can be seen, $(A \cap B) \subset S$, $(A \cap B) \subset A$, and $(A \cap B) \subset B$.

As another example, consider the universe consisting of S: {sophomores who participate in a learning experiment}. If the set A is defined by A: {students of age 20 or over} and the set C is defined by C: {students assigned to a control group}, then the intersection of the two sets consists of $A \cap C$: {students over the age of 20 assigned to a control group}.

2-7 MUTUALLY EXCLUSIVE OR DISJOINT SETS

Consider the universe S: {students at the University of California}. The set F: {freshmen} and the set J: {juniors} are both subsets of S. However, there is no student who is both a freshman and a junior at the same time. Therefore, the intersection of F and J is empty. That is, $F \cap J = \emptyset$. Since the intersection of F and J has no members, it is said that the set of freshmen and the set of juniors are mutually exclusive or disjoint sets.

Thus, if two sets A and B have no members in common, they are said to be *disjoint* or *mutually exclusive sets*. This means that their intersection is empty and is denoted by $A \cap B = \emptyset$.

A little reflection will show that complementary sets are always disjoint since the elements of A cannot be simultaneously in $\bar{A}$. Thus it is always true that $A \cap \bar{A} = \emptyset$.

2-8 THE UNION OF TWO SETS

Consider again the population of University of California students and the subsets M: {males} and F: {freshmen}. If the set of males is combined with the set of freshmen, a new set is created that contains, in addition to the freshman males, all of the sophomore males, the junior males, the senior males, the graduate males and all of the female freshmen. The appearance of the last-named set {female freshmen}

in the union of males with freshmen may come as a surprise to some students. But note that freshmen actually consist of male freshmen and female freshmen; therefore, inclusion of female freshmen in the union of males with freshmen is to be expected. This combined set U is called the union of the set of males and freshmen. Clearly, this set contains more elements than the set of males and without doubt is larger than the set of freshmen.

More formally, let S represent a universe or population. Let A be a subset of S and let B be a subset of S. The elements of S that are in A or in B or in their intersection are said to constitute the *union* of A and B. This set is denoted by $A \cup B$ As this suggests, the union of set A and its complement exhaust the universe so that $A \cup \bar{A} = S$.

For the set of {males} and {freshmen} at the University of California, it follows that $(M \cap F) \subset (M \cup F)$, $M \subset (M \cup F)$, and $F \subset (M \cup F)$. It should be noted that the union of males with freshmen contains every student of the universe except those females who are not freshmen. Thus, the complement of the union of males and freshmen is the intersection of females and nonfreshmen, or $\overline{M \cup F} = \bar{M} \cap \bar{F}$.

2-9 VENN DIAGRAMS

Sometimes an understanding of some of the properties of sets discussed in the previous sections can be facilitated by drawing diagrams called Venn diagrams, named after J. Venn, 1834–1923, who originated them. An example of a Venn diagram is shown in Figure 2-1, which is a Venn representation of a universe S and the subsets A and $\bar{A}$ contained in the universe. The area that is outside of A but still in S is the complement of A. As can be seen, $A \cap \bar{A} = \emptyset$ and $A \cup \bar{A} = S$.

In Venn diagrams, sets are associated with areas or regions. Even though areas represent continuous variables or sets with an uncountable infinity of values, they can be used symbolically to represent discrete or qualitative sets. For example, in Figure 2-1, S might refer to the universe of humans living in the United States. A could be the set of males with $\bar{A}$ referring to the set of females. In a similar fashion,

Figure 2-1. Venn diagram of complementary sets.

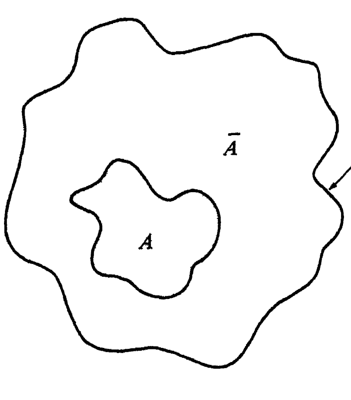
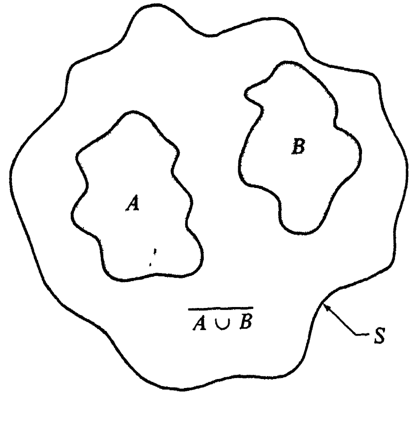


Figure 2-2. Venn diagram of disjoint sets



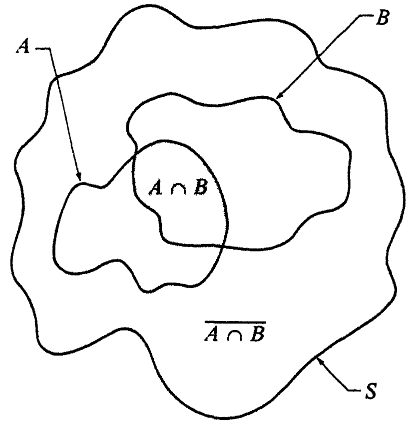


Figure 2-3. Venn diagram of the intersection of two sets

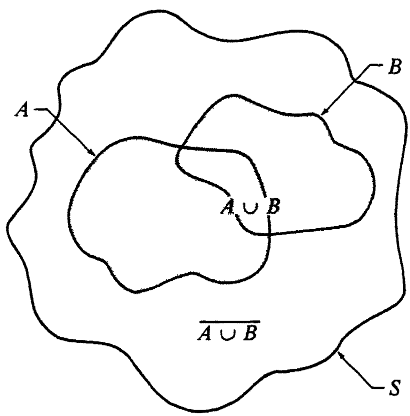


Figure 2-4. Venn diagram of the union of two sets.

S might refer to the set of outcomes when a balanced die is tossed. A could refer to the set of even outcomes, A: {2,4,6}, with $\bar{A}$ referring to the set of odd outcomes, $\bar{A}$: {1,3,5}. Finally, S might refer to the set of weights attained by people on their forty-fifth birthday with A referring to the set of weights exceeding 120 pounds and $\bar{A}$ referring to the set of weights less than or equal to 120 pounds. In set notation, $A: \{X > 120\}$ and $\bar{A}: \{X \leqslant 120\}$, where $X =$ weight in pounds of any one citizen of the United States.

Figure 2-2 is a Venn representation of a universe S and two disjoint subsets A and B contained in the universe. As can be seen, $A \cap B = \emptyset$. In this diagram, the union of A with B is represented by all of the area in A plus all of the area in B. The union of A and B, $A \cup B$, consists of the marked area. All that is unmarked represents the complement of A with B, $\overline{A \cup B}$.

Figure 2-3 is a Venn representation of a universe S and two intersecting subsets A and B contained in the universe. The area that is common to both A and B constitutes the intersection of A with B. As can be seen, $(A \cap B) \subset A$, $(A \cap B) \subset B$, $(A \cap B) \subset S$, and $(A \cap B) \subset (A \cup B)$. The unmarked part of S is $\overline{A \cap B}$. Figure 2-4 is a Venn representation of the union of two intersecting sets A and B contained in a universe S. The area that is either in A or in B or in their intersection constitutes the union of A with B. As can be seen, $A \subset (A \cup B)$, $B \subset (A \cup B)$, and $(A \cap B) \subset (A \cup B)$.

To further illustrate the notion of sets, consider the following example. Six upper-division students in a study on tonal acuity were given a test in music appreciation. Their grades in increasing numerical order are given by the following set of numbers: S: {70, 73, 81, 91, 99, 100}. The scores obtained by the four males are given by M: {70, 73, 99, 100}. The scores obtained by the two juniors are given by J: {81, 100}.

1. The scores of the females are given by the set $\bar{M}$: {81,91}.
2. The scores of the seniors are given by the set $\bar{J}$: {70,73,91,99}.
3. The scores made by the males or the juniors are given by $M \cup J$: {70, 73, 81, 99, 100}.
4. The scores made by the male juniors are given by $M \cap J$: {100}.
5. The scores made by the students who are not male juniors are given by $\overline{M \cap J}$: {70,73,81,91,99}, which should be recognized as being the same set as the union of the nonmales with the nonjuniors since $\bar{M} \cup \bar{J}$: {70,73,81,91,99}. This equivalence of sets illustrates a fundamental property of sets that is summarized in a theorem of set algebra and is popularly referred to as one of deMorgan's laws of complementation. This law says that the complement of the intersection of two sets is equal to the union of the complements of the two sets. Symbolically, $\overline{M \cap J} = \bar{M} \cup \bar{J}$.
6. The scores made by the female seniors are given by $\bar{M} \cap \bar{J}$: {91}, which should also be recognized as being the same as $\overline{M \cup J}$ since this set contains the one element {91}. This set-equivalence illustrates another of deMorgan's laws of complementation. This second law says that the complement of the union of two sets is equal to the intersection of the complements of the two sets. Symbolically, $\overline{M \cup J} = \bar{M} \cap \bar{J}$.
7. The scores made by the male seniors are given by $M \cap \bar{J}$: {70,73,99}.
8. The scores made by the female juniors are given by $\bar{M} \cap J$: {81}.
9. Finally, one should note that $(M \cap J) \cup (M \cap \bar{J}) \cup (\bar{M} \cap J) \cup (\bar{M} \cap \bar{J}) = S$. This last property is illustrated in Figure 2-5.

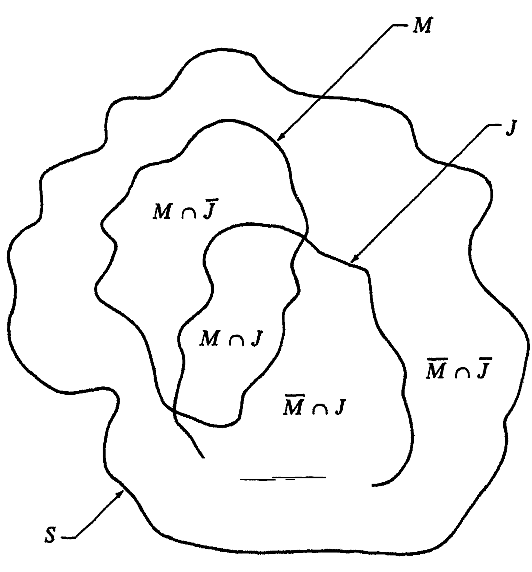
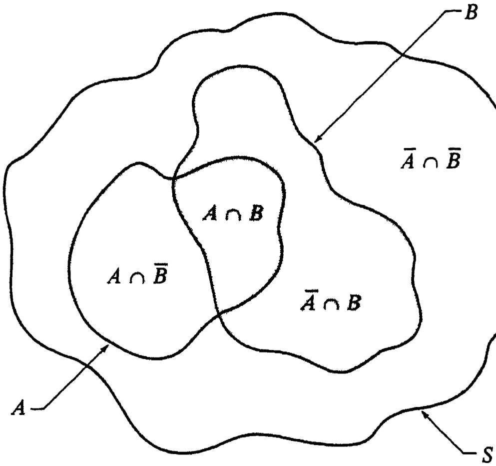


Figure 2-5. Venn diagram of the separation of a universe into four disjoint subsets.



	A	$\bar{A}$
B	$A \cap B$	$\bar{A} \cap B$
$\bar{B}$	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$

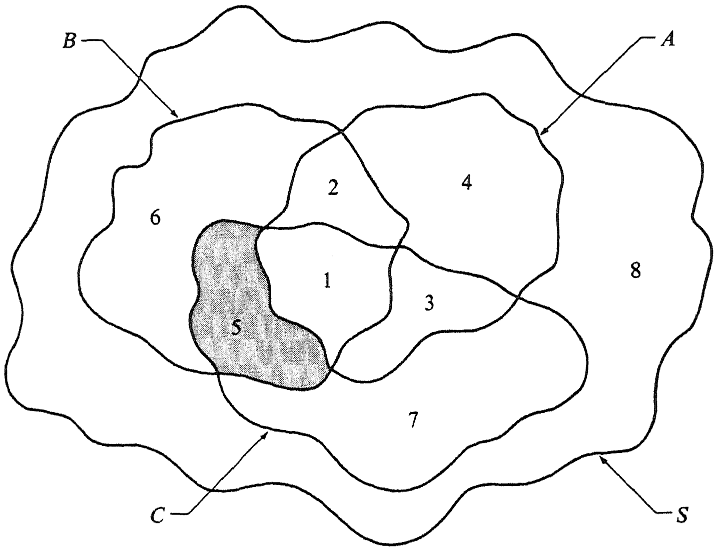
Figure 2-6. Venn diagram as a 2×2 table

2-10 SET REPRESENTATION BY TABLES

As the previous example indicates, two intersecting subsets divide the universe into four mutually exclusive or disjoint sets. Not only may these four sets be illustrated in Venn diagram form but they may also be shown as the cells of a 2×2 table. This is the usual or more popular way for presenting set relations in scientific journal articles. The correspondence between the Venn-diagram representation and the table summary of the intersection of two sets A and B is shown in Figure 2-6. Corresponding sets are shown by similar markings.

A Venn-diagram representation is somewhat confusing for a discussion involving the intersection of three sets. In this case, tables are much easier to read and generally easier to discuss. When three sets intersect, eight different mutually exclusive subsets are produced. These are illustrated in Venn-diagram form and in two different tabled forms in Figure 2-7. For example, set 5 refers to the set $\bar{A} \cap B \cap C$.

Consider S: {one hundred women who submit to illegal abortions}. They can be divided into C: {Caucasians} and $\bar{C}$: {non-Caucasians}; A: {under 21 years of age} and $\bar{A}$: {21 or older}; and R: {strong religious training} and $\bar{R}$: {weak religious training}. One way to summarize the results for these one hundred women would be as shown in Figure 2-8. The set $C \cap \bar{A} \cap \bar{R}$: {white women with weak religious training and of age 21 or over} is represented by the marked cell of the table.



Form one for the intersection of three sets

	A		$\bar{A}$	
	B	$\bar{B}$	B	$\bar{B}$
C	1	3	5	7
$\bar{C}$	2	4	6	8

Form two for the intersection of three sets

A				$\bar{A}$			
B		$\bar{B}$		B		$\bar{B}$	
C	$\bar{C}$	C	$\bar{C}$	C	$\bar{C}$	C	$\bar{C}$
1	2	3	4	5	6	7	8

Figure 2-7. Venn diagram as a 2 × 2 × 2 table for the intersection of three sets.

Figure 2-8. Table of subsets for three dichotomous classifications.

Religious training	*Caucasians*		*Non-Caucasians*		
	UNDER 21	21 AND OVER	UNDER 21	21 AND OVER	TOTAL
Strong					
Weak					
TOTAL					100

2-11 SETS AS EVENTS

Consider the set of numbers that could be generated by the tosses of a balanced six-sided die. These outcomes are given by the elements of the set S: {1,2,3,4,5,6}. Suppose a die were to be tossed and suppose a face with 6 spots appeared. One could say that the element of the set S given by "$X=6$" has been observed or else one could use contemporary statistical terminology and say that the *event* "$X=6$" was observed. In a like manner if a face with 2 spots appeared, one could say that the *event* "$X=2$" has occurred.

As another example, suppose a coin is to be tossed twice in succession. Possible ordered outcomes are given by S: {(H,H), (H,T), (T,H), (T,T)}, where T represents a tail and H represents a head. If the event "two tails" occurs, it is known immediately that the simple outcome or event A: {(T,T)} has occurred. If the event exactly "one head" has occurred, it is known that one of the complex elements of the set B:{(H,T), (T,H)} has occurred. In this case, it is impossible to state which one. In like manner, if it is known that in the event at least one head has occurred, it is known that one of the elements of the set C: {(H,H), (H,T), (T,H)} has occurred. Thus, to say that an event has occurred is the same as saying that an element has been observed from the set that describes the event.

As another example, the set A: {2,4,6} consists of the even outcomes when a balanced die is tossed. If a die is tossed and if "$X=4$" one can say either that the event "$X=4$" or the event "even number of spots" has occurred. In the first case, "$X=4$" is referred to as one of the six "simple" events of the universe. The event "even number of spots" consists of the collection of the three simple events "$X=2$," "$X=4$," and "$X=6$." Thus, to say that an even number appeared is the same as saying that an element of A: {2,4,6} has been observed.

The simple events of S correspond to the individual distinct elements of the universe. Their union exhausts the universe. From the simple events, compound events can be generated by collecting the simple events of S that have a particular property. For the dice problem, the simple events are given by "$X=1$," "$X=2$," "$X=3$," "$X=4$," "$X=5$," and "$X=6$." The compound event "even number" consists of the simple events "$X=2$," "$X=4$," and "$X=6$." As these examples suggest, if an element of the set A occurs, then the event A has also occurred.

As a further example, consider the population of women who submit to illegal abortions. If a woman of this population is selected to be interviewed and if she is a Negro, it can be said that the event "Negro" has occurred, meaning of course, that a member of the set N: {Negro} has been selected.

2-12 THE NUMBER OF ELEMENTS OF A FINITE SET

When a set is finite, it is possible to count the number of elements it contains. While most of the sets encountered in the behavioral sciences do not have this property, it is instructive to consider such sets since they help in understanding some of the elementary concepts of probability theory that are basic to modern statistical methods. With this in mind, denote the number of elements in a universe with a

finite number of elements by the symbol $N(S)$. In the school district with the 29,857 students, $N(S) = 29{,}857$. The girls, G, of this school district constitute a subset of S, that is, $G \subset S$. Let the number of girls in the subset G be denoted by $n(G)$. If the school district has 15,095 girls, $n(G) = 15{,}095$. Clearly, the complement of $G = \bar{G} = B$ so that $n(B) = 29{,}857 - 15{,}095 = 14{,}762$.

In terms of this notation:

1. If $A \subset S$, then $n(A) < N(S)$ and if $A \subseteq S$, then $n(A) \leqslant N(S)$.
2. If A and $\bar{A}$ are complementary sets, then they jointly exhaust S, that is, $n(A) + n(\bar{A}) = N(S)$.
3. If A and B are mutually exclusive subsets, then $n(A \cup B) = n(A) + n(B)$.
4. If A and B are not mutually exclusive subsets, then $n(A \cup B) = n(A) + n(B) - n(A \cap B)$, since the elements in the intersection are counted twice and must be subtracted once.
5. If A and B are disjoint, then $n(A \cap B) = 0$.

Consider a deck of 52 playing cards. Let K refer to the set of kings and let C refer to the set of clubs. Clearly, $n(K) = 4$ and $n(C) = 13$. Since there is a king that is also a club, $n(K \cap C) = 1$, and therefore the number of cards in the union of the kings with the clubs is shown by $n(K \cup C) = n(K) + n(C) - n(K \cap C) = 4 + 13 - 1 = 16$. The 16 distinct cards consist of the 13 clubs and the remaining three kings.

Furthermore, if Q consists of the set of queens, then $n(Q) = 4$. Since the set of queens and the set of kings have no cards in common, $n(Q \cap K) = 0$, and therefore the number of cards in the union of the kings with the queens is given by $n(Q \cup K) = n(Q) + n(K) - n(Q \cap K) = 4 + 4 - 0 = 8$. The eight distinct cards consist of the four kings and the four queens.

2-13 THE CARTESIAN PRODUCT OF SETS—SAMPLE SPACES

Suppose one were to take a coin and toss it. Possible outcomes are either heads or tails, which can be denoted in set notation as $S_1: \{H,T\}$. Now, if the coin were tossed a second time, the complete set of outcomes for the second toss could be denoted as $S_2: \{H,T\}$. Suppose one wished to speak of the set of outcomes when a coin is tossed twice. To do this one would have to list all ordered pairs of outcomes, which in this case would be given by $S_1 \otimes S_2: \{(H,H), (H,T), (T,H), (T,T)\}$. The set of such ordered pairs is called the *cartesian product* of S_1 and S_2 and is denoted by $S_1 \otimes S_2$. The cartesian product symbol, $\otimes$, should not be confused with the "times" symbol $\times$ that indicates the ordinary arithmetic operation of multiplication.

It is worth noting that the cartesian product of two sets is itself a set in which each element consists of an ordered pair of elements determined by taking each element of S_1 as it appears in its roster and pairing it with all the elements of S_2 as they appear in their roster.

If the coin is now tossed a third time, then the complete list of ordered triplets would be determined by taking the cartesian product of $S_1 \otimes S_2$ with S_3, which

gives $(S_1 \otimes S_2) \otimes S_3$: {$(H,H,H)$, (H,T,H), $(T,H,H,)$, (T,T,H), (H,H,T), (H,T,T), (T,H,T), (T,T,T)}.

In a more general form, suppose A is a finite set with elements A: $\{a_1,a_2, \ldots, a_K\}$ and suppose B is a finite set with elements B: $\{b_1,b_2, \ldots, b_L\}$ The *cartesian product* of A with B is defined to consist of all ordered pairs of every element of A combined with every element of B. Thus, the cartesian product of A with B is given by

$$A \otimes B: \left\{ \begin{matrix} (a_1,b_1), (a_1,b_2), \ldots, (a_1,b_L) \\ (a_2,b_1), (a_2,b_2), \ldots, (a_2,b_L) \\ \cdots\cdots\cdots\cdots\cdots \\ (a_K,b_1), (a_K,b_2), \ldots, (a_K,b_L) \end{matrix} \right\}$$

If $n(A) = K$ and $n(B) = L$, then $n(A \otimes B) = KL$ As an example of the use of the cartesian product in education research, consider a classroom with 10 students who are to be given a five-word spelling test. For any one of the children, the sets of possible numbers of correctly spelled words is given by S: {0,1,2,3,4,5}. Consider the first two children listed in the teacher's grade book. The complete listing of possible pairs of scores for these two children is given by

$$S_1 \otimes S_2: \left\{ \begin{matrix} (0,0), (0,1), (0,2), (0,3), (0,4), (0,5) \\ (1,0), (1,1), (1,2), (1,3), (1,4), (1,5) \\ (2,0), (2,1), (2,2), (2,3), (2,4), (2,5) \\ (3,0), (3,1), (3,2), (3,3), (3,4), (3,5) \\ (4,0), (4,1), (4,2), (4,3), (4,4), (4,5) \\ (5,0), (5,1), (5,2), (5,3), (5,4), (5,5) \end{matrix} \right\}$$

For the 10 children given the test the complete set of possible outcomes is given by $S_1 \otimes S_2 \otimes S_3 \otimes S_4 \otimes S_5 \otimes S_6 \otimes S_7 \otimes S_8 \otimes S_9 \otimes S_{10}$, where each S_i is given by S_i: {0,1,2,3,4,5}. This set is given by

$$\left\{ \begin{matrix} (0,0,0,0,0,0,0,0,0,0), (0,0,0,0,0,0,0,0,0,1), \ldots, \\ (4,5,5,5,5,5,5,5,5,5), \quad (5,5,5,5,5,5,5,5,5,5) \end{matrix} \right\}$$

Clearly, this is an extremely large set with $6 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6 = 6^{10} = 60{,}466{,}176$ elements or tuples of 10 numbers.

The set of elements generated by the cartesian product of sets is called by statisticians the *sample space*. It consists of the entire set of possible outcomes that can be observed in a particular context or experiment. Each individual tuple of numbers or classes is called an event or outcome of the experiment. The term *experiment* is used in an extremely broad sense by statisticians. For example, the tossing of a coin is referred to as an experiment. The possible outcomes of this experiment are heads or tails. The ordered tossing of one coin or the simultaneous tossing of two coins is also referred to as an experiment in which the set of outcomes is given by S: {(H,H), (T,H), (H,T), (T,T)}, where (H,H) refers to the event that a head appeared on the first toss and a head appeared on the second toss or that heads appeared

simultaneously on both coins, depending upon whether the discussion is about the tossing of one coin two times or the tossing of two coins simultaneously.

With respect to the experiment consisting of the five-word spelling test given to 10 students, the complete set of outcomes has been indicated in the cartesian product of $S_1 \otimes S_2 \otimes \cdots \otimes S_{10}$. The experiment produces one of those elements.

To further illustrate the notion of a sample space, consider 10 subjects assigned to participate in a learning experiment in which they are told to associate 15 nonsense words with 15 everyday objects in as few trials as possible. Suppose, furthermore, that half the students are required to recite the alphabet backwards between each trial so as to reduce the amount of intertrial rehearsal. If the number of trials for the control subjects to learn the list is reported first and the number of trials for the experimental subjects is reported second, the cartesian product or sample space for this particular study will be as follows: $(S_1 \otimes S_2 \otimes S_3 \otimes S_4 \otimes S_5) \otimes (S_6 \otimes S_7 \otimes S_8 \otimes S_9 \otimes S_{10})$: $\{[(1,1,1,1,1), (1,1,1,1,1)], [(1,1,1,1,2), (1,1,1,1,1)], \ldots\}$.

In this case it would be impossible to state the last element of the set since the number of trials needed for the slowest person to learn the list is unknown. Most likely, the maximum number of trials would be associated with one of the experimental subjects, since it is reasonable to expect the recitation of the alphabet backwards between trials to interfere with the learning process; theoretically, however, the complete tabulation could be accomplished.

As one might suspect, behavioral scientists generally do not write out the cartesian product for every experiment or study they conduct, mainly because it is time consuming and, as will be shown later, unnecessary. However, the notion of a cartesian product is essential to the development of elementary probability theory. Later it will be used to introduce the notion of a random variable and to explain elementary ideas of sampling variability and probability distribution.

2-14 SUMMARY

While a population generally consists of a collection of objects, it is convenient to associate a population or universe with all possible numerical values or exhaustive classes that a certain variable can assume over the collection of objects under study. These collections are the universes of investigation in behavioral research. The mathematical counterpart of a population is called a set. A set is a simple aggregate or collection of well-defined objects of any sort. To say that a set is well defined is to imply that one should have no problem in deciding whether or not a specified object is or is not an element of a set being considered. Sets are described by listing the elements on a roster, by verbally describing the set elements, or by specifying a range of values the objects of the set may assume.

A set that has no members is called an empty set. Two sets are said to be equal if they contain identical elements. If the elements of one set are contained in a second set, the first set is said to be a subset of the second set. The elements of the universe that are not contained in a subset of the universe are said to constitute the complementary set of the originally constructed set. The elements that are common to

two sets are said to constitute the intersection of the two sets. If two sets have no elements in common they are said to be disjoint or mutually exclusive. The set of elements that belong to two, not necessarily disjoint, sets constitutes the union of the two sets. Set relations can be illustrated by Venn diagrams or by table representations. Most journal articles use tables, since they are easier to interpret.

An important use of sets in statistics is found in the cartesian product of sets. The cartesian product of two sets is determined by taking the elements of one set and pairing them with each element of the second set. The set of ordered pairs so generated is called the cartesian product of the two sets.

EXERCISES

Exercises and questions that are starred are either based on discussions or tables in the text or related to other problems in the same or following chapters.

2-1. Suggest examples of populations in behavioral research that

(*a*) Can be studied by direct measurement.
(*b*) Can be studied only by indirect measurement.
(*c*) Can never be studied, at least with the techniques so far developed by man.

2-2. Which of the following are finite sets, countable infinite sets, and uncountable infinite sets? Give reasons for your answers.

(*a*) Possible scores on the Stanford-Binet intelligence test.
(*b*) The grains of sand along the Pacific Coast of the United States.
(*c*) The galaxies of the universe.
(*d*) Children in elementary school who received a minimally balanced diet at home during the past school year.
(*e*) Ages at death of American men who smoke.
(*f*) The percentage of correct answers to a 50-item multiple-choice test.
(*g*) Alcoholic consumption of skid-row inhabitants during the month of December in San Francisco, California.
(*h*) Patients in state mental hospitals who are diagnosed as schizophrenic on the first day of a new year.
(*i*) Graduating high school students granted scholarships to attend college this year.
(*j*) The total weights of passengers on scheduled airline trips between San Francisco and New York City.

2-3. A college sophomore class in psychology consists of 400 students Fifty of these students are to be selected to participate in an experiment that consists of a control and an experimental condition. Within the universe are a number of subsets defined by:

A: {students over 20 years of age}
B: {students who earned a B grade on the first midterm}
C: {students assigned to the control condition in the psychological experiment}
D: {students who dropped out of the class before the final}
E: {students assigned to the experimental condition in the psychological experiment}

Define the following sets. Are any of them empty sets?

(*a*) $A \cup C$
(*b*) $A \cup E$
(*c*) $A \cup (C \cup E)$
(*d*) $\overline{C \cup E}$
(*e*) $A \cap B$
(*f*) $B \cap \bar{D}$
(*g*) $B \cap (C \cup E)$
(*h*) $C \cap E$
(*i*) $\bar{D} \cap (C \cup E)$
(*j*) $\overline{(\bar{B} \cup D)}$

2-4. Represent the sets of Exercise 2-3 as Venn diagrams.

***2-5.** The table shows the statistics gathered in a hypothetical study of illegal abortions of unwed mothers. How many women are members of the following subsets?

(*a*) Caucasian women.
(*b*) Caucasian women with strong religious training over age 21.
(*c*) Women with strong religious training or women over age 21.
(*d*) Women who are non-Caucasian and over 21 or women who are Caucasian and under 21.
(*e*) Women with strong religious training over age 21 who are either Caucasian or non-Caucasian.

Religious training and age	*Caucasian mother*		*Non-Caucasian mother*		*Total*
	UNDER 21	OVER 21	UNDER 21	OVER 21	
Strong	20	18	19	6	63
Weak	36	13	23	11	83
Total	56	31	42	17	146

2-6. Construct the cartesian product for the two sets:

Experimental condition: {control, treatment}
Sex of subjects: {male, female}

2-7. A true-false test consists of four items. Construct the sample space for the complete set of possible outcomes. In how many of the outcomes is the total number of true answers greater than 2?

2-8. In a study on school integration, a random sample of adults was asked:

(*a*) Do you agree that for the best education of our students elementary schools should be integrated? Strongly agree______ Moderately agree______ Moderately disagree______ Strongly disagree______

(*b*) Do you agree that for the best education of our students junior high schools should be integrated? Strongly agree______ Moderately agree______ Moderately disagree______ Strongly disagree______

Construct the cartesian product for the two sets of possible responses. Show what the cartesian product looks like when presented as a table for a report or manuscript

***2-9.** In a blindfold testing experiment, subjects are asked to identify four specified brands of cola drinks. The set of the number of correctly identified brands is given by X: {0,1,2,4} Explain why 3 is not an element of this set.

***2-10.** In the experiment of Exercise 2-9, two experts in the identification of cola drinks were tested. Construct the cartesian product for their joint set of outcomes. What is the possible set of values for their joint total of correctly identified colas; i.e., what is the set of outcomes for $T = X_1 + X_2$ where

X_1 = number of correctly identified colas by the first subject tested
X_2 = number of correctly identified colas by the second subject tested

3 INTRODUCTION TO PROBABILITY THEORY

CASINOS VS. THE COMPUTERS

Cannes, France The betting of many European croupiers is that gambling will be as dead as the dinosaurus within 20 years.

Signor Bertolini, manager of the Municipal Casino in San Remo, denies it, but Italian tourists insist that he chased a computerized German gang out of his cream-puff stucco casino last month after it had won 40 million lire.

The group was using the Richard Jareki system · Watch the same croupier at the same roulette table for several days. Note down the winning numbers until 30,000 are collected.

Feed these into a computer, and learn the five that have won most regularly. Then play them for all you are worth

By Ferris Hartman, Chronicle Foreign Service. By permission, from San Francisco Chronicle, San Francisco, California, September 24, 1966.

3-1 THE PROBABILITY OF AN EVENT FOR EQUALLY LIKELY OUTCOMES

In Chapter 2, the association between algebraic sets and statistical events was discussed. Furthermore, the number of elements in a set (finite) was also presented. With these two concepts, it is quite easy to develop a theory of probability that has wide application in behavioral research. For this development, it is convenient to start with a definition of probability for equally likely events or outcomes. Consider a finite universe S and a subset A contained in S. Let the number of elements in S be denoted by $N(S)$ and let the number of elements in A be denoted by $n(A)$. If all elements of S have an equal chance of occurring under any random selection of elements, then the *probability of A relative to S* is defined to be the ratio of $n(A)$ to $N(S)$, and one writes

$$P(A) = \frac{n(A)}{N(S)}$$

$P(A)$ is called the probability of the set A or the event A.

As an example of the use of this definition, consider a perfectly balanced die with uniform density and with faces of equal size and symmetry. For such a die, the set of outcomes is given by $S: \{1,2,3,4,5,6\}$. If the die is taken in hand, is vigorously shaken, and then thrown against a flat wall, every face has an equal chance of appearing. Consider the set of outcomes in which the uppermost face has 3 spots showing. Denote this event by A. Since only one element of S has this property, $n(A) = 1$. Since $N(S) = 6$, it follows that the probability of throwing a 3 is given by $P(A) = n(A)/N(S) = \frac{1}{6}$. Consider the set of even-numbered faces and denote this set by B. The set B consists of B: $\{2,4,6\}$ and $n(B) = 3$, so that the probability of an even-numbered face is given by $P(B) = n(B)/N(S) = \frac{3}{6} = \frac{1}{2}$.

It should be noted that the words *chance* and *random* have appeared in the definition of the probability of a set or event. These terms will remain undefined in this book. The definition of these terms has engaged, without success, the greatest of mathematicians, and to attempt a definition here could lead the discussion of probability off into many tangents not directly related to the goals of this book. As might be expected, there are many ways to define these terms; however, the definitions that are finally agreed upon must not lead to embarrassing paradoxes and inconsistencies. To avoid these pitfalls the reader is asked to rely on intuition and past experiences to help make the concept of probability meaningful.

It must be emphasized that this introduction to probability has been restricted to finite sets with equally likely outcomes. This is an extremely limiting presentation. It is being used here since it makes it easy to present the ideas of probability. However, for completeness a more encompassing and general probability model will be presented. As will be shown, the model of equally likely outcomes is simply a special case of the more general theory.

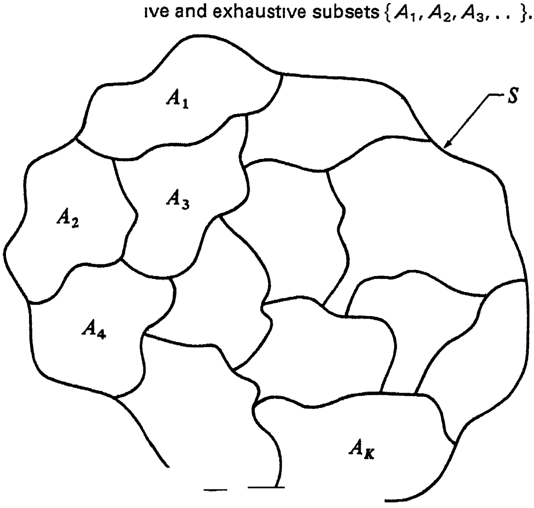
3-2 GENERAL PROBABILITY MODEL

The probability model presented and discussed in Section 3-1 has been restricted to a very special case that has limited usefulness for research problems of the behavioral sciences. For this special case it has been assumed that all possible events in the sample space have the same probability of occurrence. This means that each element in the sample space has an equal probability of occurrence on any one trial or selection. Since this requirement restricts the application of the theory to problems of the everyday real world, a more general theory is needed that incorporates the model of equally likely probabilities as a special case. For this more general probability model, let the sample space be partitioned into a set of mutually exclusive subsets, $A_1, A_2, A_3, \ldots$, as shown in Figure 3-1 The general method of partitioning may create a finite or an infinite number of sets. In Figure 3-1 the number is finite. To simplify the presentation it will be assumed that the number of subsets is finite. Denote this exhaustive set of mutually exclusive subsets as $\{A_1, A_2, \ldots, A_K\}$. Since these subsets exhaust the sample space, it must be true that their union is equal to the universe. Thus

$$A_1 \cup A_2 \cup \cdots \cup A_K = S$$

Without any loss of generality it is possible to associate a real number with each subset. Let these numbers be denoted by $\{p_1, p_2, \ldots, p_K\}$. Let the number associated with A_1 be p_1, the number associated with A_2 be p_2, and, finally, the number associated with A_K be p_K. Let these numbers be called the probabilities of the various sets. For a meaningful probability model, the numbers must be controlled and not allowed to vary without limits. In particular, these numbers must be restricted so

Figure 3-1. Partition of a universe S into mutually exclusive and exhaustive subsets $\{A_1, A_2, A_3, \ldots\}$.



that each is greater than or equal to zero but less than or equal to unity. Thus, a typical p_k must satisfy the relationship $\{0 \leqslant p_k \leqslant 1\}$. As a further restriction upon these numbers, it is required that their sum be unity:

$$p_1 + p_2 + \cdots + p_k + \cdots + p_K = 1$$

Starting with these definitions and restrictions, one can generate a complete theory or probability. While this theory may be mathematically and logically satisfying, it has the extra utilitarian property that it facilitates the study of events in the real world that are of interest to behavioral scientists.

Perhaps it is useful to point out the similarities that an axiomatic study of probability has to an axiomatic development of geometry. Formal euclidean geometry has little relationship to the real world of objects and distances even though observations of the real world may have influenced its inception and development. Geometry is primarily an abstract system relating hypothetical constructs by definitions and theorems. That it has been found to be useful in describing the real world as it is observed is incidental. The points, lines, and figures of geometry are pure abstractions. They are true hypothetical constructs and are without properties of reality. The same thing is true of the numbers called probabilities. They, too, are abstractions and belong completely to the conceptual sphere.

For example, it is possible to talk about the probability of a head or tail without ever tossing a coin. In a purely abstract sense it can be said that the outcomes must be heads or tails and that the probabilities of these outcomes must be p_1 and p_2 such that $p_1 + p_2 = 1$. In like manner, the experiment that consists of tossing a die once can be conceived entirely in an abstract or conceptual sense. The set of possible outcomes could be denoted by a symbol S. Since the observed outcome would be the number of spots appearing on the upper face of the die, the set of possible outcomes could be denoted by the numbers S: $\{1,2,3,4,5,6\}$, and the probabilities of these outcomes could be denoted by P: $\{p_1,p_2,p_3,p_4,p_5,p_6\}$ such that $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$.

It should be emphasized that nothing in the theory of probability says that these possible outcomes must be equally likely. It may be useful at certain times to assume this, even though in the entire history of the world there has probably never been a die manufactured by man for which the probability of the appearance of a 1 is equal to the probability of a 2, etc. Though it is useful, at times, to make the assumption that outcomes are equally likely, it is only a convention that simplifies discussion. Thus, for the purposes of exposition, it will generally be assumed *before* a die is tossed that the probability of observing any one number from 1 to 6 showing up is equal to $\frac{1}{6}$. This will be done mainly to illustrate the principles of probability theory, and not because it is true.

Finally, it should be noted that the condition of equally likely outcomes is included in the general theory presented. If the outcomes are equally likely, with probability p_0, then $p_1 = p_2 = \cdots = p_k = \cdots = p_K = p_0$.

Theorem 3-1

If a finite set has K equally likely outcomes, then the probability of each outcome p_0 is given by

$$p_0 = \frac{1}{K}$$

Proof. By assumption,

$$p_1 = p_2 = \cdots = p_K = p_0$$

Substituting this common value into

$$p_1 + p_2 + \cdots + p_K = 1$$

we have

$$p_0 + p_0 + \cdots + p_0 = 1$$

so that $Kp_0 = 1$, and finally, $p_0 = 1/K$. This completes the proof.

This theorem states that when K events are equally likely, each has probability $1/K$ of occurring on any one trial. For the coin-tossing experiment, $K = 2$, so that $p_H = p_T = p_0 = \frac{1}{2}$. For the die-tossing experiment, $K = 6$, so that $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_0 = \frac{1}{6}$.

3-3 PROBABILITIES OF EVENTS IN FINITE SAMPLE SPACES

As an example based on the assumption of equally likely outcomes, consider an experiment that consists of the tossing of a die. The outcomes of the experiment can be represented by the set of numbers S: {1,2,3,4,5,6}. If the assumption of equal probability of occurrence is superimposed upon the outcomes and if X represents the observed outcome, it follows that

$$P(X=1) = P(X=2) = P(X=3) = P(X=4) = P(X=5) = P(X=6) = \tfrac{1}{6}$$

For example, before the die is tossed, the probability that the outcome will have three spots showing is given by $P(X=3) = \frac{1}{6}$. Note that this corresponds to the definition of probability presented earlier for equally likely outcomes. The number of outcomes satisfying the condition $X = 3$ is only 1, and since $N(S) = 6$, it follows from the definition of probability that $P(X=3) = n(X=3)/N(S) = \frac{1}{6}$.

Continue with the model of a fair die (all outcomes are equally likely) and consider the event that the outcome will be the appearance of an even number of dots. To determine this probability, let A be the event {even number of dots}. This event is given by A: {2,4,6} with $n(A) = 3$. By the definition of probability, $P(A) = \frac{3}{6} = \frac{1}{2} = .5$, since it has been assumed that all outcomes are equally likely to occur. If this last assumption had not been made, P (even number) $= .5$ would not be correct. That this is indeed the case is illustrated in the next example.

Suppose that the die to be tossed is a loaded die and that the $P(X=3)=.6$. It is quite obvious that the probability of an even number of dots appearing is less than .5, since $P(X=3)$, which is odd, is greater than .5. This must be the case since

$$P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)+P(X=6)=1$$

But, as is known, $P(X=3)=.6$, so that

$$P(X=1)+P(X=2)+P(X=4)+P(X=5)+P(X=6)=.4$$

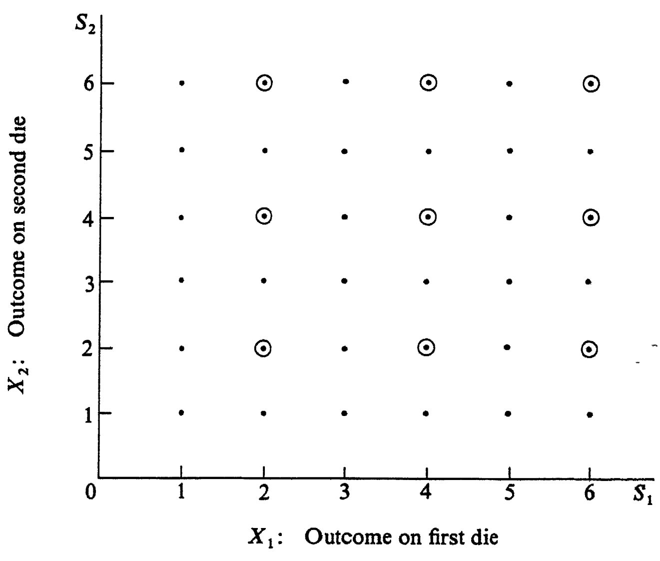
and

$$P(X=2)+P(X=4)+P(X=6)\leqslant .4$$

3-4 PROBABILITIES OF EVENTS IN A TWO-DIMENSIONAL SAMPLE SPACE

Consider an experiment that consists of the simultaneous tossing of two identical dice. If A is the event {outcome on the first die}, then the possible set of outcomes is given by $S_1:\{1,2,3,4,5,6\}$, and if B is the event {outcome on the second die}, then the possible set of outcomes for B is given by $S_2:\{1,2,3,4,5,6\}$. Consider the joint event produced by tossing the dice together. The outcomes must consist of pairs of numbers that are generated by the cartesian product $S=S_1\otimes S_2$. Let the observed outcomes be denoted by X_1 and X_2. These outcomes are the 36 ordered pairs $(x_1,x_2)\in(S_1\otimes S_2):\{(1,1),\ (1,2),\ (1,3),\dots,(x_1,x_2),\dots,(6,5),\ (6,6)\}$. This sample space is illustrated in Figure 3-2. If the restriction of equally likely outcomes is

Figure 3-2. Sample space $S=S_1\otimes S_2$ for the outcomes generated by the simultaneous tossing of two dice



placed upon each of the 36 pairs of outcomes, it follows that

$$P(X_1 = 1 \cap X_2 = 1) = P(X_1 = 1 \cap X_2 = 2) = \cdots$$
$$= P(X_1 = 6 \cap X_2 = 5) = P(X_1 = 6 \cap X_2 = 6) = \tfrac{1}{36}$$

Consider the following examples, based upon this model:

1. Let A be the event $\{X_1 = 3 \cap X_2 = 5\}$. Since there is only one point in the cartesian product that has the property A, $n(A) = 1$ and $P(A) = n(A)/N(S) = \frac{1}{36}$.
2. Let B be the event {one die shows a 3 and the other shows a 5}. In this case, the number of outcomes having this property is 2. They are: $\{(X_1 = 3 \cap X_2 = 5)$ and $(X_1 = 5 \cap X_2 = 3)\}$. Thus, $n(B) = 2$ and $P(B) = n(B)/N(S) = \frac{2}{36}$.
3. Let C be the event {the sum of the dots on the upper two faces is 8}. Five points have this property. They are $\{(X_1 = 2 \cap X_2 = 6)$, $(X_1 = 3 \cap X_2 = 5)$, $(X_1 = 4 \cap X_2 = 4)$, $(X_1 = 5 \cap X_2 = 3)$, $(X_1 = 6 \cap X_2 = 2)\}$. Thus, $n(C) = 5$ and $P(C) = n(C)/N(S) = \frac{5}{36}$.
4. Let D be the event {the number of dots on both of the dice is even}. The number of outcomes with this property is 9. They are circled in Figure 3-2. Thus, $n(D) = 9$ and $P(D) = n(D)/N(S) = \frac{9}{36}$.

Notice that as long as the events in the sample space are equally likely, it is necessary only to count the number of outcomes having the desired property and then to divide this number by the total number of possible outcomes in the universe to find the probability of interest. If the points are not equally likely, counting will be of little value. To clarify this statement, consider what would happen to the four probabilities of A, B, C, and D if $P(X_1 = 3) = .2$ and $P(X_2 = 5) = .6$. None of the above statements would be correct; most would be false, with the probabilities being too low.

The definitions of probability given by the equally likely concept or the more general model are quite workable in the handling of abstract gambling systems and games, but they find less application in the study of external environment and for this a different definition is needed. This need is satisfied through the relative-frequency definition of probability. Through this second definition, empirical research is placed on a firm foundation so that the application of statistical inference procedures can be defended on a meaningful theoretical basis.

3-5 THE RELATIVE-FREQUENCY INTERPRETATION OF PROBABILITY

The relative-frequency definition of probability is the following:

If, whenever a sequence of simple but identical experiments is conducted, the ratio of the number of times event A occurs to the total number of trials made is nearly p, and if when longer series of trials are made the ratio is usually nearer to p than it is to any other number, then it is agreed in advance to define the probability

of A as p, and to say that the probability of A is equal to the number p. Furthermore, this is symbolized by writing $P(A) = p$.

Operationally, this definition means that the probability of A is equal to the limit of the ratio $n(A)/N(S)$ as $N(S)$ increases without bounds. The actual determination of p depends upon performing an experiment that is repeated over and over again. After each performance of the experiment, the ratio of the total number of A's observed to the total number of times the experiment is performed is determined. If this ratio remains closer to the number p more than it does to any other number, the number p, by convention, is called the probability of A. For any finite number of trials, the ratio $n(A)/N(S)$ is called the relative frequency of A. If $N(S)$ increases without limit, this relative frequency converges upon the value of p.

In a certain sense the relative-frequency definition of probability stipulates that p can never be known since an infinite number of trials are required to determine its exact value. While p cannot be known exactly, the relative frequency of A in any finite number of trials can serve as an estimate of $P(A)$, which can be made as close to the true value as is desired. In Section 9-10, a method is described that can be used to determine the number of trials or experiments that must be performed to give a numerical value of $n(A)/N(S)$ that can serve as a reasonable estimate of p. The utility of the method is based on the fact that $n(A)/N(S)$ stabilizes in a relatively small number of trials. This stabilizing property of the ratio of $n(A)$ to $N(S)$ is easy to illustrate.

In Table 8-1 is presented a list of hypothetical test scores made by 500 students on a standardized test. For this test consider the event that a person selected at random has a score in the 50's, $X\colon \{50 \leqslant X \leqslant 59\}$. For the first 10 students, three have scores in the range $X\colon \{50 \leqslant X \leqslant 59\}$, so that for the first 10 trials, $n(A)/N(S) = \frac{3}{10} = .30$. Since this is only an estimate of p, a notation for this estimate that has gained near universal acceptance by statisticians is $\hat{p}$. Thus, for 10 trials, $\hat{p} = .30$. Continuing in this manner, the convergence of $\hat{p}$ to p is readily noted:

In the first 10 trials $n(A) = 3$, so that $\hat{p} = \frac{3}{10} = .30$
In the first 20 trials $n(A) = 5$, so that $\hat{p} = \frac{5}{20} = .25$
In the first 30 trials $n(A) = 8$, so that $\hat{p} = \frac{8}{30} = .27$
In the first 40 trials $n(A) = 12$, so that $\hat{p} = \frac{12}{40} = .30$
In the first 50 trials $n(A) = 18$, so that $\hat{p} = \frac{18}{50} = .36$

One could continue in this manner until the fluctuations in $\hat{p}$ were as small as desired. For the entire set of data, $p = .34$. The convergence of $\hat{p}$ to p is illustrated in Table 3-1. Notice that after 50 trials the relative frequency of scores in the 50's never deviated by more than two percentage points from the true value of .34. Thus, 10 percent of the universe produced an estimate of p that deviated no more than 2 percent from the correct value. The ability that small samples have to provide such excellent estimates of unknown quantities is a constant source of surprise and is of immense value in behavioral research.

This may suggest that 30,000 trials may be more than sufficient to estimate the probabilities of a "balanced" roulette wheel at San Remo, Italy. But when one realizes that a European roulette wheel has 37 slots with numbers extending from 0 to 36, then it is seen that 30,000 trials must be distributed over the estimation of 37 probabilities. In any case, the advice to "learn the five that have won most regularly, then play them for all you are worth" is certainly based upon the belief in the relative-frequency interpretation of probability. If physical forces on the wheel remain fairly constant, then a player following this advice should do quite well.

Frequently, in behavioral research, the probabilities of rather complex events will be required. Fortunately, these probabilities can be easily determined from the probabilities of simple events by means of an algebra for probabilities. This algebra for certain simple cases will be established in the following sections.

Table 3-1. Example illustrating the convergence of the relative frequency of $n(A)/N(S)$ to $P(A)$.

Number of trials	*Number of scores in the 50's*	*Relative frequency*
10	3	.30
20	5	.25
30	8	.27
40	12	.30
50	18	.36
60	21	.35
70	23	.33
80	27	.34
90	29	.32
100	33	.33
110	36	.33
120	40	.33
130	44	.34
140	45	32
150	51	.34
160	56	.35
170	58	34
180	64	.36
190	68	.36
200	71	.36
300	102	.34
400	132	.33
500	170	34

3-6 PROBABILITY OF THE COMPLEMENT OF A

Theorem 3-2

If the probability of A is given by $P(A)$, then the probability of $\bar{A}$ is given by $P(\bar{A}) = 1 - P(A)$.

Proof Consider a finite sample space with $N(S)$ elements. Suppose A is an event in this space whose probability of occurrence is known. That is,

$$P(A) = \frac{n(A)}{N(S)}$$

Since the elements of A and the elements of $\bar{A}$ exhaust S,

$$N(S) = n(A) + n(\bar{A})$$

Dividing both sides of this equation by $N(S)$, it follows that

$$\frac{N(S)}{N(S)} = \frac{n(A)}{N(S)} + \frac{n(\bar{A})}{N(S)}$$

or

$$1 = P(A) + P(\bar{A})$$

Thus

$$P(\bar{A}) = 1 - P(A)$$

This completes the proof.

This theorem states that the probability of the complement of A is equal to 1 minus the probability of A. As an example, consider a school with 1,000 children of which 600 are boys and 400 are girls. The relative frequency of boys is $n(B)/N(S) = \frac{600}{1000} = .60$. If this particular school were the universe of an educational study, .60 could be treated as though it were the probability of selecting a boy at random from the school. Since the complement of boy is girl, the probability of selecting a girl is $P(G) = 1 - P(\bar{G}) = 1 - P(B) = 1 - .60 = .40$. Of course, this probability could have been determined by simply computing $n(G)/N(S) = \frac{400}{1000} = .40$.

As another example, consider the tossing of a coin for which the event {heads} has probability $P(H) = \frac{1}{3}$. Since the complement of heads is tails, the probability of the event {tails} or {not heads} is $P(T) = 1 - P(H) = 1 - \frac{1}{3} = \frac{2}{3}$.

As another example, consider selecting a card from a deck of 52 playing cards. The entire deck could be represented in a two-dimensional sample space, as shown in Figure 3-3. The suits {clubs, diamonds, hearts, spades} are listed along the X axis; the face value of the cards {ace, two, three, . . . , king} are listed along the Y axis. If a card is drawn from this space, it must have two properties, a suit value and a numerical value. The probability of drawing the three of hearts is $\frac{1}{52}$, since only one card has both properties. The probability of drawing anything but the

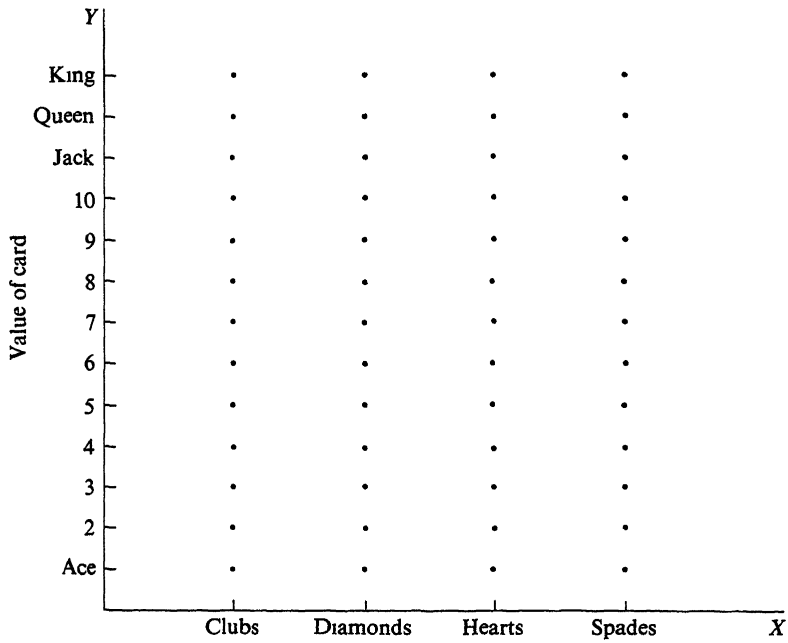


Figure 3-3. Sample space for the 52 cards of a standard bridge deck

three of hearts is $(1 - \frac{1}{52})$ or $\frac{51}{52}$. In a like manner, the probability of selecting a king is $\frac{4}{52}$. Therefore, the probability of selecting anything but a king is $(1 - \frac{4}{52})$ or $\frac{48}{52}$.

3-7 THE ADDITION RULE FOR MUTUALLY EXCLUSIVE EVENTS

Theorem 3-3

If A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.

Proof. If events A and B are mutually exclusive, it follows that their intersection is empty. Symbolically, $A \cap B = \emptyset$. Therefore, the number of elements in the union of A with B is given by

$$n(A \cup B) = n(A) + n(B)$$

If both sides of this equation are divided by $N(S)$, one obtains

$$\frac{n(A \cup B)}{N(S)} = \frac{n(A)}{N(S)} + \frac{n(B)}{N(S)}$$

from which it follows that

$$P(A \cup B) = P(A) + P(B)$$

This completes the proof.

This theorem may be extended to any number of mutually exclusive sets. For example, suppose $\{A, B, C, \ldots\}$ are mutually exclusive. Then

$$P(A \cup B \cup C \cup \cdots) = P(A) + P(B) + P(C) + \cdots$$

Selection of a king and a queen on a single drawing of a card is an example of the occurrence of two mutually exclusive events because it is impossible to choose a king and a queen at the same time. Since there are four points in the sample space that have the property of being a king and four that have the property of being a queen, the probability of choosing a king is $\frac{4}{52}$ and the probability of choosing a queen is $\frac{4}{52}$. Thus, the probability of choosing a king or a queen is

$$\begin{aligned} P(K \cup Q) &= P(K) + P(Q) \\ &= \tfrac{4}{52} + \tfrac{4}{52} \\ &= \tfrac{8}{52} \end{aligned}$$

As another example of these results, consider the simultaneous tossing of two fair dice and let the event of interest be that their sum will be 9. This is a complex event that can be broken down into a set of simple mutually exclusive events. Let X_1 equal the number of dots showing on one of the dice and let X_2 equal the number of dots showing on the remaining die. X_1 could be equal to 6 and X_2 could be equal to 3; or X_1 could be equal to 5 and X_2 could be equal to 4; or X_1 could equal 4 and X_2 could equal 5; or X_1 could equal 3 and X_2 could equal 6. A little reflection will prove that the four pairs of events are mutually exclusive. If one occurs at any one tossing, the other could not have occurred. Thus, $P(X_1 + X_2 = 9) = P(X_1 = 6 \cap X_2 = 3) + P(X_1 = 5 \cap X_2 = 4) + P(X_1 = 4 \cap X_2 = 5) + P(X_1 = 3 \cap X_2 = 6)$. Since $P(X_1 = 6 \cap X_2 = 3) = P(X_1 = 5 \cap X_2 = 4) = P(X_1 = 4 \cap X_2 = 5) = P(X_1 = 3 \cap X_2 = 6) = \frac{1}{36}$, $P(X_1 + X_2 = 9) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{4}{36} = \frac{1}{9}$. This means that chances are 1 in 9 that a total of 9 will appear when a pair of fair dice are tossed. This probability could be interpreted in terms of the relative-frequency definition of probability. It would mean that if one took these two dice and repeatedly tossed them and counted the number of times a 9 appeared and after each toss calculated the ratio of the number of times a 9 appeared to the number of times the dice were tossed, this ratio would converge eventually upon $\frac{1}{9}$. However, in this case it would take more than 1,000 tosses to obtain an estimate of p that would not deviate from $\frac{1}{9} = .1111\ldots$ by more than .02.

3-8 PROBABILITIES OF EVENTS THAT ARE NOT MUTUALLY EXCLUSIVE

Theorem 3-4

If A and B are not mutually exclusive subsets, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. As was shown in Section 2-8, the number of elements in the union of two sets that are not mutually exclusive is given by

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

Therefore the probability of the union of A and B is, by definition,

$$P(A \cup B) = \frac{n(A \cup B)}{N(S)}$$

$$= \frac{n(A) + n(B) - n(A \cap B)}{N(S)}$$

$$= \frac{n(A)}{N(S)} + \frac{n(B)}{N(S)} - \frac{n(A \cap B)}{N(S)}$$

Since $P(A) = n(A)/N(S)$, $P(B) = n(B)/N(S)$, and $P(A \cap B) = n(A \cap B)/N(S)$, it follows that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

This completes the proof.

If A and B are mutually exclusive, $A \cap B = \emptyset$ and $n(A \cap B) = 0$, and $P(A \cap B) = 0$. Thus, $P(A \cup B) = P(A) + P(B)$. This is the formula presented in Section 3-7 for adding the probabilities of mutually exclusive events.

To illustrate the use of the general formula for $P(A \cup B)$, consider selecting a card from a pack of 52 cards. What is the probability that this card is a club or is an honor card? To bridge players, honor cards consist of the tens, jacks, queens, kings, and aces of each suit Since some of the honor cards are clubs, the intersection of the set of clubs and the set of honor cards is not empty. Let C be the set of clubs and H the set of honor cards. Thus

$$P(C \cup H) = P(C) + P(H) - P(C \cap H)$$

There are 20 honor cards in the deck since each suit has 5 honor cards. Since there are 52 cards in the deck, $P(H) = \frac{20}{52}$. There are 13 clubs in the deck. The probability of drawing a club is $P(C) = \frac{13}{52}$. Five of the clubs are honor cards. Thus, the probability of drawing an honor card that is a club is equal to $P(C \cap H) = \frac{5}{52}$. Finally,

$$P(C \cup H) = \frac{13}{52} + \frac{20}{52} - \frac{5}{52} = \frac{28}{52}$$

Table 3-2. Hypothetical distribution of test success according to sex.

Outcome	*Boys*	*Girls*	*Total*
Passed	2	20	22
Failed	18	10	28
Total	20	30	50

Consider the hypothetical data of Table 3-2, which show the distribution of pass or fail on a test administered to a class of 20 boys and 30 girls. The probability that a student in this population is either a girl or has failed the test is given by

$$\begin{aligned} P(G \cup F) &= P(G) + P(F) - P(G \cap F) \\ &= \tfrac{30}{50} + \tfrac{28}{50} - \tfrac{10}{50} \\ &= \tfrac{48}{50} \end{aligned}$$

Note that this probability could also have been determined by using one of de Morgan's laws of complementation. For this approach,

$$\begin{aligned} P(G \cup F) &= 1 - P(\overline{\bar{G} \cup \bar{F}}) \\ &= 1 - P(\bar{G} \cap \bar{F}) \\ &= 1 - P(B \cap P) \\ &= 1 - \tfrac{2}{50} \\ &= \tfrac{48}{50} \end{aligned}$$

3-9 STATISTICAL INDEPENDENCE OF TWO EVENTS

Consider two students who are to be given the spelling test described in Section 2-13. The variable to be observed on each student is the number of words correctly spelled on the test. Possible scores for the first student are $X_1: \{0,1,2,3,4,5\}$; possible scores for the second student are the same and are given by $X_2: \{0,1,2,3,4,5\}$. The sample space for the simultaneous scores of these two students consists of the 36 pairs of numbers in the matrix on page 31. If these students do not copy from one another and if they take the test so that their final scores are in no way related to one another, then it is said that the outcomes are independent. While independence of performance on a test is essential for evaluating the performance of each child, it is also important for the *valid* use of the commonly employed statistical methods. The definition of statistical independence of two events is customarily stated in probability terms and is, in reality, a probability statement about the intersection of two sets. The definition of *statistical independence* is as follows:

Two events, A and B, are statistically independent if the probability of their intersection, $A \cap B$, is equal to the product of the probabilities of their separate events. Thus, A and B are independent if

$$P(A \cap B) = P(A)P(B)$$

To illustrate the concept with a simple example, suppose a card is selected from a deck of 52 cards. After its face value is observed, the card is returned to the deck and another card is drawn. What is the probability that a king is selected on the first draw and a queen is selected on the second draw? Let K_1 denote that a king is selected first and Q_2 denote that a queen is selected on the second trial. On the

first drawing, the deck contains four kings. Thus, $P(K_1) = \frac{4}{52} = \frac{1}{13}$. On the second drawing, the deck has been returned to its original state and therefore it contains four queens. Thus, $P(Q_2) = \frac{4}{52} = \frac{1}{13}$. These two events are said to be statistically independent since what happened on the first draw does not affect or influence what happens on the second draw. Thus $P(K_1 \cap Q_2) = P(K_1)P(Q_2) = (\frac{1}{13})(\frac{1}{13}) = \frac{1}{169}$. If this experiment were repeated 169 times, the number of times one would select a king followed by a queen is about 1. If it were repeated 1,690 times, the number of times one would expect the king to be followed by the queen is about 10.

The probability of this complex event could have been determined by direct application of the definition of probability as the ratio of $n(A \cap B)/N(S)$. However, to use the definition it would be necessary to count points in the sample space generated by the experiment of selecting 2 cards from a deck of 52 cards. Reliance on the definition for the determination of the probability of interest would quickly prove that short-cut methods are needed to simplify the computation of probabilities. Knowledge of the independence of events is one way to achieve this simplification.

In the above example, the appearance of king and queen are *ordered*. If the ordering is ignored, the probability of drawing a king and a queen is not equal to $\frac{1}{169}$. The event of interest could happen in two mutually exclusive ways: king first, then a queen; or queen first, then a king. Therefore

$$\begin{aligned} P(K \cap Q) &= P[(K_1 \cap Q_2) \cup (Q_1 \cap K_2)] \\ &= P(K_1 \cap Q_2) + P(Q_1 \cap K_2) \end{aligned}$$

Since the events of drawing a king and drawing a queen are independent,

$$\begin{aligned} P(K \cap Q) &= P(K_1)P(Q_2) + P(Q_1)P(K_2) \\ &= (\tfrac{1}{13})(\tfrac{1}{13}) + (\tfrac{1}{13})(\tfrac{1}{13}) \\ &= \tfrac{2}{169} \end{aligned}$$

As another example, consider tossing a coin three times. If the coin is tossed three times in succession, the probability of a head on the first toss equals the probability of a head on the second toss, which in turn equals the probability of a head on the third toss. If X represents the number of heads in three tosses, the probability that X is equal to 3 is given by

$$P(X = 3) = P(H_1 \cap H_2 \cap H_3)$$

Since the outcomes on each toss are unrelated to the outcomes on the remaining tosses, it follows that the tosses are statistically independent and $P(H_1 \cap H_2 \cap H_3) = P(H_1)P(H_2)P(H_3)$. If the coin is fair, heads and tails are equally likely, so that $P(H_1) = P(H_2) = P(H_3) = \frac{1}{2}$. Thus,

$$P(X = 3) = (\tfrac{1}{2})(\tfrac{1}{2})(\tfrac{1}{2}) = \tfrac{1}{8}$$

Consider the event two heads or $\{X = 2\}$. This can occur in three mutually exclusive ways. They are $\{(H_1 \cap H_2 \cap T_3),\ (H_1 \cap T_2 \cap H_3),\ (T_1 \cap H_2 \cap H_3)\}$. If

one of these events occurs, it would be impossible for either of the other two to occur; as a result, the intersection of the first event with the other two is empty, and the theorem on addition of probabilities of mutually exclusive events may be applied. Thus

$$P(X=2)=P(H_1 \cap H_2 \cap T_3)+P(H_1 \cap T_2 \cap H_3)+P(T_1 \cap H_2 \cap H_3)$$

Since each of the tosses is independent,

$$P(X=2)=P(H_1)P(H_2)P(T_3)+P(H_1)P(T_2)P(H_3)+P(T_1)P(H_2)P(H_3)$$

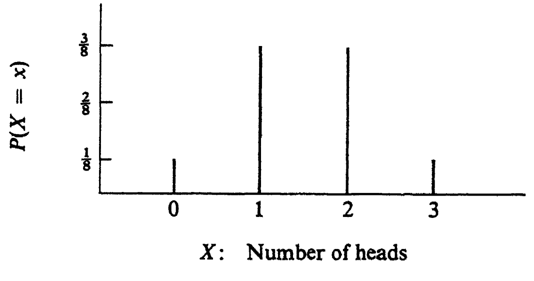
Since $P(H_1)=P(H_2)=P(H_3)=\frac{1}{2}$, it follows that

$$\begin{aligned} P(X=2) &= (\tfrac{1}{2})(\tfrac{1}{2})(\tfrac{1}{2})+(\tfrac{1}{2})(\tfrac{1}{2})(\tfrac{1}{2})+(\tfrac{1}{2})(\tfrac{1}{2})(\tfrac{1}{2}) \\ &= \tfrac{1}{8}+\tfrac{1}{8}+\tfrac{1}{8}=\tfrac{3}{8} \end{aligned}$$

Consider the event 1 head or $\{X=1\}$. This is the same as the event 2 tails. Since $P(H)=P(T)=\frac{1}{2}$, the probability of two tails should equal the probability of two heads Thus, $P(X=1)=\frac{3}{8}$. The probability of obtaining three heads in three tosses is the same as no heads in three tosses. Thus, $P(X=3)=P(X=0)=\frac{1}{8}$. Since the events $\{(X=3), (X=2), (X=1), (X=0)\}$ exhaust all possible outcomes of the universe, it must be that $P[(X=3) \cup (X=2) \cup (X=1) \cup (X=0)]=1$. Since these events are mutually exclusive, their probabilities add. Thus, $P(X=3) + P(X=2)+P(X=1)+P(X=0)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=\frac{8}{8}=1$.

It is often useful to represent probabilities of mutually exclusive and exhaustive events graphically. Along the X axis, possible outcomes $\{0,1,2,3\}$ are listed. Along the Y axis, a line equal in height to the probability of X is drawn. The probabilities for the coin-tossing example are shown in Figure 3-4. The complete graph represents the probability distribution for the number of heads that appear when a fair coin is tossed three times. Later it will be seen that probability distributions play a major role in statistical theory, and their extremely close connection to behavioral research will be emphasized repeatedly in later sections of this book.

Figure 3-4. Probability distribution for all possible outcomes for the tossing of a fair coin three times.



3-10 CONDITIONAL PROBABILITY

One of the most frequently occurring questions of behavioral research is whether variables, usually qualitative, are independent or related. In answering questions of this nature, the definition of statistical independence is used. However, the same questions can be answered by using an easier to understand, but related, concept. This related concept is called conditional probability. As will be seen in the next section, statistical independence and conditional probability are concepts intimately associated with one another.

Before conditional probability is defined, consider an unpublished study by Marascuilo (1965) concerning the attitudes of junior high school students toward a general school environment. Negro students at two different schools in an urban community were asked how well they liked school during the present year as compared to the previous school year. Their responses were as shown in Table 3-3.

Table 3-3. Attitudes of Negro students at two different schools concerning how well they like school during the present year as compared to the previous school year.

Response	*School A*	*School B*	*Total*
More	157	39	196
Less	60	64	124
Total	217	103	320

For these students, the liking of school is clearly related to the school attended. At School A, $\frac{157}{217}$ or .72 like school more, while at School B, the corresponding figure is $\frac{39}{103}$ or .38. If the school classification is ignored, the proportion of students who like school more is $\frac{196}{320}$ or .61. From these figures, it would be concluded that there has been a more favorable change in attitude toward school at School A than at School B.

The two probabilities associated with the two schools are called conditional probabilities. The probability figure that ignores schools is called the unconditional or marginal probability.

When the conditional probabilities differ from the unconditional probability a notation is required to differentiate between them. The notation generally employed in the literature is as follows:

1. The unconditional probability of liking school more is simply
$P(\text{more}) = \frac{196}{320} = .61$

2. The conditional probability of liking school more, given that the student is enrolled in School A, is
$P(\text{more}|\text{School } A) = \frac{157}{217} = .72$
3. The conditional probability of liking school more, given that the student is enrolled in School B, is
$P(\text{more}|\text{School } B) = \frac{39}{103} = .38$

Upon first exposure, many students frequently have problems differentiating between unconditional and conditional probabilities. Much of the problem can be reduced by realizing that both concepts are probabilities. Unconditional probabilities are determined in terms of S, the universe, while conditional probabilities are determined from subsets of S. Thus, when discussing conditional probabilities, the set or basis of reference has changed to one of the subsets of the universe.

As will be seen, conditional probabilities play a major role in statistical methods and theory. In fact, it might be said that all research is devoted to estimating conditional probabilities and testing hypotheses to determine whether conditional probabilities are equal to unconditional probabilities.

For a general discussion, think of a universe whose elements may be classified on two dimensions. Let these dimensions be represented by A and B. All elements may be classified as being $(A \cap B)$ or $(A \cap \bar{B})$ or $(\bar{A} \cap B)$ or $(\bar{A} \cap \bar{B})$. The number of elements associated with each of these mutually exclusive and exhaustive subsets of S can be represented as shown in Table 3-4.

Table 3-4. Typical four-fold table for the joint classification of a universe according to two characteristics.

Characteristic	A	$\bar{A}$	*Total*
B	$n(A \cap B)$	$n(\bar{A} \cap B)$	$n(B)$
$\bar{B}$	$n(A \cap \bar{B})$	$n(\bar{A} \cap \bar{B})$	$n(\bar{B})$
Total	$n(A)$	$n(\bar{A})$	$N(S)$

If an element is taken from the universe, what is the probability that the element has properties A and B at the same time? By definition, the probability of A intersected with B is given by the ratio of the number of elements having the property A and B to the number of elements in the sample space. Thus

$$P(A \cap B) = \frac{n(A \cap B)}{N(S)}$$

This in turn is the same as

$$P(A \cap B) = \frac{n(A \cap B)}{N(S)} \frac{n(A)}{n(A)} = \frac{n(A)}{N(S)} \frac{n(A \cap B)}{n(A)}$$

The first fraction is, by definition, the probability of A. The second fraction is the ratio of the number of events having both A and B to those that are A. This is the conditional probability of B, given that A has occurred. In the previous example, this is simply the probability of liking school more, given that the student is at School A. It is the ratio of the number of more responses at School A to the total number of students at School A. Using the notation of conditional probability, we have

$$P(A \cap B) = P(A)P(B|A)$$

This last equation can always be used to determine the probability of the intersection of two events or it can be used to determine the conditional probability of one event, given that the other event has occurred

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

As an example of the use of this result, suppose an experiment consists of giving an examination to a group of freshmen. Before presentation of the test one can classify the freshmen according to their sex, boys and girls. After giving the examination, one can further classify the students according to whether or not they passed or failed. The possible outcomes are: boys who pass ($B \cap P$); boys who fail ($B \cap F$); girls who pass ($G \cap P$); and girls who fail ($G \cap F$). The basic question of this experiment can be stated as: "Is the passing of this examination independent of sex?" "Is the probability of passing this test equal for the two sexes?" Or, "Is the probability of passing this test higher for boys than it is for girls?"

Suppose that after the test is given the distribution by sex of pass and fail is as shown in Table 3-2. From the figures reported in Table 3-2, the probability of passing, independent of sex, is given by

$$P(P) = \frac{n(P)}{N(S)} = \frac{22}{50} = .44$$

The probability of a boy passing is given by

$$P(P|B) = \frac{n(P \cap B)}{n(B)} = \frac{2}{20} = .10$$

The probability of a girl passing is given by

$$P(P|G) = \frac{n(P \cap G)}{n(G)} = \frac{20}{30} = .67$$

Girls have a greater probability of passing the examination than do boys. This suggests that passing the examination is associated with or dependent upon sex. If sex made no difference, then it should be true that

$$P(P|B)=P(P|G)=P(P)$$

This suggests a simple test for determining whether or not two variables are independent. If the variables are independent, the conditional probabilities will equal the unconditional probability. In this example, the conditional probabilities do not equal the marginal probability: $P(P|B)\neq P(P|G)\neq P(P)$. Consequently, passing the examination is not independent of sex.

Since passing is not independent of sex, it follows that $P(P \cap B)\neq P(P)P(B)$. That this is, indeed, true is easy to demonstrate. By definition,

$$P(P \cap B)=\frac{n(P \cap B)}{N(S)}=\frac{2}{50}=.04$$

$$P(P)=\frac{n(P)}{N(S)}=\frac{22}{50}=.44$$

$$P(B)=\frac{n(B)}{N(S)}=\frac{20}{50}=.40$$

Since $P(P)P(B)=(.44)(.40)=.1760$ is not anywhere near $P(P \cap B)=.04$, it must follow that success on the test is not statistically independent of sex. This suggests a close connection between statistical independence and conditional probabilities. The connection is stated as follows: If A and B are statistically independent, then

1. $P(A|B)=P(A|\bar{B})=P(A)$
2. $P(A \cap B)=P(A)P(B)$

In Chapter 16, it will be seen that form (1) is used in what is termed the two-sample binomial problem, while form (2) will be used in Chapter 17 in what is termed the one-sample correlation problem. Until these methods are described, either one of these relations may be used to test for statistical independence.

It should also be noted that if two variables are statistically independent, then all of the following relationships hold:

$$P(A)=P(A|B)=P(A|\bar{B}) \qquad P(B)=P(B|A)=P(B|\bar{A})$$
$$P(\bar{A})=P(\bar{A}|B)=P(\bar{A}|\bar{B}) \qquad P(\bar{B})=P(\bar{B}|A)=P(\bar{B}|\bar{A})$$

Although any one of these relationships may be used to test for independence, in practice researchers ordinarily use the first pair of each set. If these relationships are satisfied, then $P(A \cap B)=P(A)P(B)$ is also true.

The general relationship $P(A \cap B)=P(A|B)P(B)$ is always true, but the statement $P(A \cap B)=P(A)P(B)$ is true only if the variables A and B are statistically

independent. Suppose now that A and B are independent. Then $P(A \cap B) = P(A)P(B) = P(A)P(B|A)$ and therefore, $P(B) = P(B|A)$ This means that the occurrence of the event B is unrelated to the occurrence of the event A.

In Section 3-9 two events were said to be statistically independent if $P(A \cap B) = P(A)P(B)$. In the notation of a four-fold table, the product of $P(A)$ and $P(B)$ can be expressed as $[n(A)/N(S)]\,[n(B)/N(S)]$. By definition, the probability of the joint event A and B equals $n(A \cap B)/N(S)$. If A and B are independent, then these two different ways for computing the $P(A \cap B)$ should be equal Thus, if

$$\frac{n(A \cap B)}{N(S)} = \frac{n(A)}{N(S)}\frac{n(B)}{N(S)}$$

then A and B are independent. The data in a four-fold table can be used to make a second test of the possible independence of two classificatory variables A and B. For this test, compute $P(A) \times P(B)$ and compare this product with $P(A \cap B)$. If the two quantities are equal, the variables are independent; if unequal, the variables are dependent.

In the same study in which Negroes were asked about liking school, Caucasian students were also polled. The data for the Caucasian students were as shown in Table 3-5. As before, one would like to know whether attitude is independent of

Table 3-5. Attitudes of Caucasian students at two different schools concerning how well they like school during the present year as compared to the previous school year.

Response	*School A*	*School B*	*Total*
More	189	63	252
Less	65	58	123
Total	254	121	375

the school attended. In other words, does the student's school membership affect his attitude toward school? Both methods for testing independence will be illustrated. For the first method,

1. $P(\text{more}) = \frac{252}{375} = .67$
2. $P(\text{more}|\text{School } A) = \frac{189}{254} = .74$
3. $P(\text{more}|\text{School } B) = \frac{63}{121} = .52$

When the conditional probabilities are compared with the marginal probability, the attitude differences between schools are apparent. If school membership and attitude toward school were independent, then it would be true that

$$P(\text{more}) = P(\text{more}|\text{School } A) = P(\text{more}|\text{School } B)$$

In behavioral research, population data are generally unavailable, so that an investigator must rely on the conclusions he can draw from samples. While relationships may be of a specified nature in the universe of study, the corresponding relationships are not ordinarily identical or similar in the sample even in those cases where the variables are truly independent in the population. Because of this, one needs a decision rule to aid in deciding on the basis of sample data whether or not variables in the population are statistically independent. Until exact statistical decision-rule methodology is developed and presented, it is suggested that one temporarily use the following rule of thumb to decide about possible independence of variables. If the average of the positive deviations of $P(A|B)$ and $P(A|\bar{B})$ from $P(A)$ does not exceed .10, one should behave as though A and B are statistically independent. Applying this rule to the previous example for the Caucasians, we find that

$$|P(M) - P(M|A)| = |.67 - .74| = .07$$
$$|P(M) - P(M|B)| = |.67 - .52| = .15$$

The average of the positive deviations is .11. Since $.11 > .10$, this *quick test* would lead to the conclusion that school membership and attitude are not independent. Note, the conclusion is that attitude and school membership are not independent. The conclusion is *not* that school membership causes differences in attitude, even though it may be true.

For the second method, data in the four-fold table can be used to compute the following probabilities:

$$P(\text{more}) = \tfrac{252}{375} = .67$$
$$P(\text{School } A) = \tfrac{254}{375} = .68$$
$$P(\text{more}) \times P(\text{School } A) = (.67)(.68) = .4556$$

From the table, one reads directly that

$$P(\text{more and School } A) = \tfrac{189}{375} = .50$$

Since $.4556 \neq .50$, there is some evidence that attitude is related to school.

As has been stated, experimental data do not always satisfy the relationships defined by statistical independence even though in the population the relationship is nearly or exactly true. The reason for this is that chance and random effects tend to disrupt the ideal conditions of statistical theory. Nature is unaware of definitions and rules and as a result, experimental outcomes deviate somewhat from theory. Thus in practice one concludes that in the universe of study, A and B are statistically independent if $P(A \cap B)$ and $P(A) \times P(B)$ are approximately equal in the sample.

As another example, suppose that the hypothetical statistics reported in Table 3-6 were obtained in a study of 600 voters' attitudes toward recalling a local school board. One would like to determine whether there is a difference in attitude toward recall on the basis of political affiliation. For these data,

Table 3-6. Attitudes toward recall of school board by party affiliation for a set of hypothetical statistics.

Attitude	*Republican*	*Democrat*	*Independent*	*Total*
For recall	120	90	40	250
Against recall	180	110	60	350
Total	300	200	100	600

$$P(\text{for recall}) = \tfrac{250}{600} = .417$$
$$P(\text{for recall}\,|\,\text{Republican}) = \tfrac{120}{300} = .400$$
$$P(\text{for recall}\,|\,\text{Democrat}) = \tfrac{90}{200} = .450$$
$$P(\text{for recall}\,|\,\text{Independent}) = \tfrac{40}{100} = .400$$

For these data the average absolute deviation is given by

$$\frac{|.400 - .417| + |.450 - .417| + |.400 - .417|}{3} = \frac{.017 + .033 + .017}{3} = .022$$

Since this average does not exceed .10, attitude toward recall and party affiliation can be treated as being statistically independent variables. It is reasonable to consider differences of the magnitude of .017 and .033 as chance deviations from .000. If the study were to be repeated with another sample of 600 voters, it would be reasonable to expect the numbers and the percentages in the 2×3 table to be similar although not identical to those shown in Table 3-6.

3-11 INTRODUCTION TO GENERAL SAMPLING THEORY

Consider a classroom with B boys and G girls from which a sample of two subjects is to be selected for a psychological experiment. So as not to bias the experiment, it would not be desirable to show any individual preference for experimental subjects Thus, selection should be made so that each student in the classroom has an equal chance of being in the experiment. If this is the case, what is the probability that both subjects will be boys, both subjects will be girls, or that one subject will be a boy and the other will be a girl? To answer these questions, let X equal the number of boys in the sample. Possible values for X are $\{0,1,2\}$. Consider the event $\{X = 2\}$. One way to satisfy this event is to select a boy, return him to the class, and then select another boy, with the possibility that the same boy might be chosen as the second subject. In general, this would not be acceptable to most behavioral scientists. Such sampling is referred to as sampling with replacement. If the first subject selected is not returned to the population, the sampling scheme is referred to as sampling without replacement. As might be expected, the probabilities associated with these two types of sampling are different. An example will illustrate this point.

3-12 SAMPLING WITHOUT REPLACEMENT

Consider the probability of choosing two boys from a classroom consisting of B boys and G girls. This is the same as the probability of selecting a boy on the first selection and a different boy on the second selection. These selections are not independent, since if a boy is chosen on the first sampling, the probability of choosing a boy on the second selection decreases. Thus, $P(X=2)=P(\text{boy on first trial and boy on second trial})$. Letting B_1 be the first event and B_2 the second event, we have

$$P(X=2)=P(B_1 \cap B_2)=P(B_1)P(B_2|B_1)$$

Thus, the probability of interest is given by the probability of selecting a boy on the first trial times the probability of choosing a boy on the second trial under the condition that a boy had been chosen on the first trial. The probability of choosing a boy on the first trial is given by

$$P(B)=\frac{n(\text{boys in class})}{N(\text{class size})}=\frac{B}{B+G}$$

The probability of choosing a second boy under the condition that the first selection produced a boy is given by

$$P(B_2|B_1)=\frac{n(\text{boys in class after one is removed})}{N(\text{remaining class size})}=\frac{B-1}{(B-1)+G}$$

Thus, the probability of interest is given by:

$$P(X=2)=\frac{B}{B+G}\frac{B-1}{B+G-1}$$

Consider the event $\{X=1\}$ or that one boy is included in the sample. This means that one subject is a boy and the second subject is a girl. This can happen in two ways: it could be that the first subject drawn is a boy and the second subject drawn is a girl or that the first subject drawn is a girl and the second subject drawn is a boy. If this is the case,

$$\begin{aligned} P(X=1) &= P\begin{bmatrix}\text{(boy on first drawing and girl on second drawing) or} \\ \text{(girl on first drawing and boy on second drawing)}\end{bmatrix} \\ &= P[(B_1 \cap G_2) \text{ or } (G_1 \cap B_2)] \end{aligned}$$

These pairs of events are mutually exclusive and cannot occur simultaneously. Therefore, by the addition rule for mutually exclusive events,

$$P(X=1)=P(B_1 \cap G_2)+P(G_1 \cap B_2)$$

Employing the same argument as that used in finding $P(X=2)$, we have

$$P(X=1)=P(B_1)P(G_2|B_1)+P(G_1)P(B_2|G_1)$$

As before, $P(B_1)=B/(B+G)$, and for the same argument, $P(G_1)=G/(B+G)$.

The probability of choosing a girl on the second trial given that the first trial produces a boy is, by definition,

$$P(G_2|B_1) = \frac{n(\text{girls in class after removing boy})}{N(\text{remaining class size})}$$

$$= \frac{G}{(B-1)+G}$$

In like manner,

$$P(B_2|G_1) = \frac{B}{B+(G-1)}$$

Thus

$$P(X=1) = \frac{B}{B+G}\frac{G}{(B-1)+G} + \frac{G}{B+G}\frac{B}{B+(G-1)}$$

$$= \frac{2BG}{(B+G)(B+(G-1))}$$

Finally, consider the event $\{X=0\}$, or that no boys are in the sample. Since the sample space can be partitioned into three mutually exclusive and exhaustive subsets—samples with two boys, samples with one boy, and samples with zero boys—it follows that

$$P(X=2) + P(X=1) + P(X=0) = 1$$

Thus

$$P(X=0) = 1 - P(X=2) - P(X=1)$$

$$= 1 - \frac{B}{B+G}\frac{B-1}{B+G-1} - \frac{2BG}{(B+G)(B+G-1)}$$

This can be simplified by placing the algebraic quantities over a common denominator. Thus

$$P(X=0) = \frac{1(B+G)(B+G-1) - B(B-1) - 2BG}{(B+G)(B+G-1)}$$

$$= \frac{B^2 + BG - B + BG + G^2 - G - B^2 + B - 2BG}{(B+G)(B+G-1)}$$

$$= \frac{G^2 - G}{(B+G)(B+G-1)}$$

$$= \frac{G(G-1)}{(B+G)(B+G-1)}$$

$$= \frac{G}{B+G}\frac{G-1}{B+G-1}$$

This could also have been determined by using the same argument as used for two boys. In essence, the selection of two girls is similar to the selection of two boys. In order to determine this probability, one need only change each B to a G and each G to a B.

Suppose there are 20 boys and 12 girls in the class. The probability of choosing two boys is given by

$$P(X=2) = \frac{B}{B+G}\frac{B-1}{B+G-1} = \left(\frac{20}{32}\right)\left(\frac{19}{31}\right) = .38$$

In other words, the probability is slightly over $\frac{1}{3}$ that two boys will be chosen. The probability of drawing a boy and a girl is given by

$$P(X=1) = \frac{2BG}{(B+G)(B+G-1)} = \frac{(2)(20)(12)}{(32)(31)} = .48$$

According to these results, the probability is slightly less than $\frac{1}{2}$ that one boy and one girl will be selected at random. Thus, there is about a fifty-fifty chance of drawing a boy and a girl, which is somewhat surprising since the ratio of boys to girls is 20 to 12. The probability of drawing zero boys is given by

$$\begin{aligned} P(X=0) &= 1 - P(X=2) - P(X=1) \\ &= 1 - .38 - .48 \\ &= .14 \end{aligned}$$

which indicates that the probability of selecting two girls at random is slightly less than $\frac{1}{6}$.

3-13 SAMPLING WITH REPLACEMENT

Reconsider the sampling problem of Section 3-12, but this time let the sampling be with replacement. This means that once a student is drawn out of the class, he or she is returned and is faced with the risk of being redrawn into the experiment. Consider the probability of choosing two boys, under the condition that the selected subject is returned to the classroom. In this case, the events are clearly independent because returning a student to the classroom means that the classroom is returned to its original state and therefore what happens on the first drawing in no way influences what happens on the second drawing This also means that the probabilities do not change but remain constant over both drawings. Since events are independent,

$$P(X=2) = P(B_1 \cap B_2) = P(B_1)P(B_2)$$

and since

$$P(B_1) = P(B_2) = P(B) = \frac{B}{B+G}$$

it follows that

$$P(X=2)=\frac{B}{B+G}\frac{B}{B+G}=\left(\frac{B}{B+G}\right)^2=p^2$$

if one lets $p=B/(B+G)$. In like manner,

$$P(X=0)=P(G_1 \cap G_2)=P(G_1)\,P(G_2)=\frac{G}{B+G}\frac{G}{B+G}=\left(\frac{G}{B+G}\right)^2=q^2$$

if one lets $q=1-p=1-B/(B+G)=G/(B+G)$.

Consider selecting one boy and one girl. As before, this can occur in two ways. Either the boy is drawn first or the girl is drawn first. Thus

$$P(X=1)=P[(B_1 \cap G_2) \text{ or } (G_1 \cap B_2)]$$

Since these events are mutually exclusive,

$$P(X=1)=P(B_1 \cap G_2)+P(G_1 \cap B_2)$$

Furthermore, since the events within the parentheses are independent,

$$\begin{aligned} P(X=1) &= P(B)P(G)+P(G)P(B) \\ &= 2P(B)P(G) \\ &= 2\frac{B}{B+G}\frac{G}{B+G}=2pq \end{aligned}$$

If the number of boys is 20 and the number of girls is 12, then these probabilities are given by

$$P(X=2)=\left(\frac{B}{B+G}\right)^2=\left(\frac{20}{32}\right)^2=\frac{25}{64}=.39$$

$$P(X=1)=2\frac{B}{B+G}\frac{G}{B+G}=2\left(\frac{20}{32}\right)\left(\frac{12}{32}\right)=\frac{30}{64}=.47$$

$$P(X=0)=\left(\frac{G}{B+G}\right)^2=\left(\frac{12}{32}\right)^2=\frac{9}{64}=.14$$

Clearly, these probability formulas are simpler to derive and use than those for sampling without replacement. Because of this, these formulas will be examined in greater detail and conditions will be stated when sampling without replacement can be treated as though it were sampling with replacement. For the applications encountered in the behavioral sciences, these probabilities will be quite close to one another in numerical value.

3-14 SUMMARY

In this chapter, probabilities for equally likely events were used to lay the foundations for statistical inference theory. It was stated that if $A \subset S$, and if the elements in S are equally likely, then the probability of A is defined as

$$P(A) = \frac{n(A)}{N(S)}$$

This was followed by a general probability model in which the probabilities of the subsets $\{A_1, A_2, A_3, \ldots\}$ were assumed to be the numbers $\{p_1, p_2, p_3, \ldots\}$ such that each p_k satisfied the inequality $\{0 \leqslant p_k \leqslant 1\}$ and such that $p_1 + p_2 + p_3 + \cdots = 1$. When the p's are equal, the general model reduces to the equally likely model.

While the general theory of probability is quite useful for studying gambling systems and games, it has less utility in the study of behavioral phenomena, since the p_k are generally unknown. To circumvent this difficulty, the relative-frequency concept of probability is called upon. According to this model, an experiment is repeated a number of times in succession and the appearance of A or $\bar{A}$ is noted. Following each trial of the experiment, $\hat{p} = n(A)/N(S)$ is determined. Over repeated trials, $\hat{p}$ fluctuates, but as $N(S)$ gets exceptionally large, the fluctuations in $\hat{p}$ approach 0 and eventually $\hat{p}$ converges upon p, the probability of A. For the most part, the validity of common statistical procedures is dependent upon this probability convergence.

Probabilities of complex events are frequently needed in behavioral studies. If S is a universe with subsets A and B with $P(A)$ and $P(B)$, then

1. The probability of the complement of A is given by
   $P(\bar{A}) = 1 - P(A)$
2. The probability of the union of A and B is given by
   $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. If A and B are mutually exclusive, then
   $P(A \cup B) = P(A) + P(B)$
4. If A and B are statistically independent, then
   $P(A \cap B) = P(A)P(B)$

One of the most important probability notions of behavioral research is that of conditional probability. Unconditional probabilities are measured relative to the universe S of a study. Conditional probabilities are measured relative to subsets in S. The most frequently asked question of conditional probabilities and of behavioral research is "Are conditional probabilities that are defined on two or more subsets of a behavioral research study equal?" If the answer is yes, one says that the characteristic being observed on the elements of the universe is statistically independent of the subset to which the elements belong. If the answer is no, one says that the variables under investigation are dependent, associated, or correlated; i.e., the characteristic or variable under investigation varies with the subsets of the universe so that knowledge of the subsets gives information concerning the characteristic being studied.

If $P(A|B)$ denotes the probability of A given the occurrence of B, and $P(A|\bar{B})$ represents the probability of A on the occurrence of B, then in terms of conditional probabilities

5. The probability of A intersected with B is given by,
   $P(A \cap B) = P(A)P(B|A)$
6. If A and B are statistically independent,
   $P(A|B) = P(A|\bar{B}) = P(A)$

Finally, an introduction to random sampling was presented for sampling with replacement and sampling without replacement. When sampling without replacement, outcomes are not statistically independent since the universe size changes as elements are removed. In sampling with replacement, outcomes are statistically independent since after each selection the universe is returned to its original state by placing the selected element back into the population pool. While sampling without replacement is the preferred sampling scheme for behavioral studies, the probability considerations for its use are not simple. Since sampling with replacement is easier to handle mathematically and since the probabilities of sampling with replacement are frequently excellent approximations to the idealized method of sampling, it will be given further considerations in the following pages of this book.

EXERCISES

3-1. Take a coin and toss it a number of times, and after each toss determine the ratio of the number of heads observed to the total number of tosses taken. How many tosses are necessary before the ratio does not deviate by more than 02 over 10 consecutive trials? (Note that the number of trials will vary with each coin and each experimenter.)

3-2. In the two-dimensional sample space shown in Figure 3-2, determine the following probabilities:

(*a*) $P(X_1 = 4 \cap X_2 = 2)$
(*b*) $P(X_1 = 4)$
(*c*) $P(X_2 = 2)$
(*d*) $P(X_1 = 4 \cup X_2 = 2)$
(*e*) $P(X_1 + X_2 = 6)$
(*f*) $P(X_1 + X_2 = 8)$
(*g*) $P[(X_1 + X_2 = 6) \cap (X_1 + X_2 = 8)]$
(*h*) $P[(X_1 + X_2 = 6) \cup (X_1 + X_2 = 8)]$

3-3. Explain how one can determine the probability that a male student at the University of California is over 6 feet tall without being compelled to measure the heights of all the male students.

3-4. Explain how one can determine the average age of incoming patients at a medical clinic.

3-5. In a study in which two methods for the teaching of third-grade reading were being tested, students were given the same achievement test. The results of the testing were as shown in the following table:

Achievement on test	*Method one Control* (C)	*Method two Experimental* (E)	*Total*
Above grade level (A)	12	22	34
Below grade level (B)	11	7	18
Total	23	29	52

Assuming that this sample is large enough to give reliable estimates of probabilities, estimate the following and answer the following questions:

(*a*) $P(A \cup E)$
(*b*) $P(\overline{A \cup E})$
(*c*) $P(A|C)$
(*d*) $P(A|E)$
(*e*) $P(A)$
(*f*) Is performance on the test independent of treatment? Why?
(*g*) $P[(B \cup E) \cap (B \cup C)]$
(*h*) $P[(B \cup E) \cup (B \cup C)]$
(*i*) Does $P(A \cap E) = P(A)P(E)$?
(*j*) What does this suggest?

***3-6.** The sets A and B are mutually exclusive but $P(A) = .6$ and $P(B) = .3$. Find the following probabilities:

(*a*) $P(\bar{A})$
(*b*) $P(A \cup B)$
(*c*) $P(A \cap B)$
(*d*) $P(A|B)$
(*e*) $P(A|\bar{B})$
(*f*) $P(\bar{A} \cap \bar{B})$
(*g*) $P(B|\bar{A})$
(*h*) $P(B|A)$
(*i*) $P(\overline{A \cup B})$
(*j*) $P(\overline{A \cap B})$

***3-7.** The sets A and B are independent but $P(A) = .6$ and $P(B) = .3$. Find the probabilities of Exercise 3-6.

***3-8.** A classroom consists of 12 boys and 20 girls:

(*a*) If a sample of two children is selected without replacement, what is the probability that the second child is a girl?

(*b*) If a sample of three children is selected without replacement, what is the probability that the third child is a girl?

(*c*) If a sample of three children is selected without replacement, what is the probability that the second and third child are girls?

(*d*) If a sample of three children is selected without replacement, what is the probability that two of the children are girls?

(*e*) Explain why the probabilities of (*c*) and (*d*) are different.

***3-9.** Estimate the probabilities of the sets in Exercise 2-5.

***3-10.** In the study on school integration in Exercise 2-8, the responses of 100 individuals to the two questions were as shown in the following table.

Attitudes toward integration of junior high schools	*Attitudes toward integration of elementary schools*				*Total*
	STRONGLY AGREE	MODERATELY AGREE	MODERATELY DISAGREE	STRONGLY DISAGREE	
Strongly agree	5	8	12	10	35
Moderately agree	6	10	8	14	38
Moderately disagree	4	2	8	2	16
Strongly disagree	1	2	1	7	11
Total	16	22	29	33	100

Assuming that the sample is large enough so that the relative frequencies are good estimators for the probability, find:

(*a*) The probability that people have identical views on both questions.

(*b*) The probability that people tend to agree more about integrating junior high schools than they do about elementary schools.

(*c*) The probability that people agree about integrating junior high schools but disagree about integrating elementary schools.

(*d*) The probability that people have completely opposite views on the two propositions.

(*e*) Explain why (*a*) and (*d*) are not complementary sets.

4 INTRODUCTION TO COMBINATORIAL ANALYSIS

. . These ideas were presented some time ago to a number of children in kindergarten. . .

It was raining and the children were asked how many raindrops would fall on New York. The highest answer was 100. They had never counted higher than 100 and what they meant to imply when they used that number was merely something very, very big—as big as they could imagine. They were asked how many raindrops hit the roof, and how many hit New York, and how many single raindrops hit all of New York in 24 hours. They soon got a notion of the bigness of these numbers even though they did not know the symbols for them. They were certain in a little while that the number of raindrops was a great deal bigger than a hundred They were asked to think of the number of grains of sand on the beach at Coney Island and decided that the number of grains of sand and the number of raindrops were about the same. But the important thing is that they realized that the number was *finite, not infinite* In this respect they showed their distinct superiority over many scientists who to this day use the word infinite when they mean some big number, like a billion billion. . . .

From *New Names for Old,* Copyright © 1940 by Edward Kasner and James Newman in *Mathematics and the Imagination,* Simon and Schuster, New York, 1940.

4-1 FUNDAMENTALS OF COUNTING

While counting the number of raindrops that could fall on New York City during a rainstorm is perhaps an impossible task, consider the simple problem of a family living in San Francisco planning a vacation in Chicago but passing through Salt Lake City both going and returning. Suppose after much family deliberation it is decided that two different routes between San Francisco and Salt Lake City appear to be interesting and worthy of travel for sight-seeing and vacation potentials and that once Salt Lake City is reached, three different and interesting routes to Chicago are available. Since there are two routes leading to Salt Lake City from San Francisco, and since there are three different routes leading from Salt Lake City to Chicago, the total number of different routes between San Francisco and Chicago going through Salt Lake City is six. These routes are shown in Figure 4-1. If one so desired, one could list the complete set of routes by forming the cartesian product of the two sets of routes.

Essentially this example illustrates what is called the fundamental principle of counting. This principle is as follows:

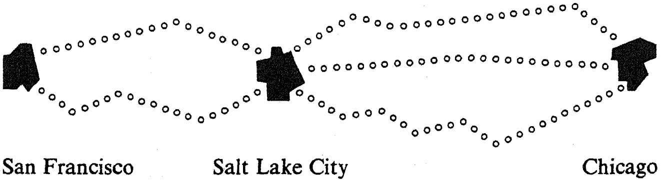
If event A_1 can occur in $n(A_1)$ ways and event A_2 in $n(A_2)$ ways, then their joint event $A_1 \otimes A_2$ can occur in $n(A_1)n(A_2)$ ways.

As might be expected, this principle of counting can be extended. If the event A_1 can occur in $n(A_1)$ ways, the event A_2 in $n(A_2)$ ways, the event A_3 in $n(A_3)$ ways, $\ldots$, and the event A_K in $n(A_K)$ ways, then the event $A_1 \otimes A_2 \otimes A_3 \otimes \cdots \otimes A_K$ can occur in $n(A_1)n(A_2)n(A_3) \cdots n(A_K)$ ways.

Thus, for the family planning its vacation, the number of different return trips passing through Salt Lake City is given by $(2)(1) = 2$, so that the number of different round trips, passing through Salt Lake City both coming and going, is $(6)(2) = 12$.

Sometimes a quick counting method can be generated by what is called a tree diagram or branching process. For the complete set of round trips for the family of the previous example, the resulting tree diagram or branching process is shown in Figure 4-2. In the figure, the first set of arrows represents the number of different routes from San Francisco to Salt Lake City. The second set of arrows represents the number of different routes from Salt Lake City to Chicago. The total number of different routes is clearly 6. In the next sets of arrows, the return trips are shown. The total number of different round trips is 12.

Figure 4-1. Vacation routes between San Francisco and Chicago that pass through Salt Lake City, for the hypothetical family.



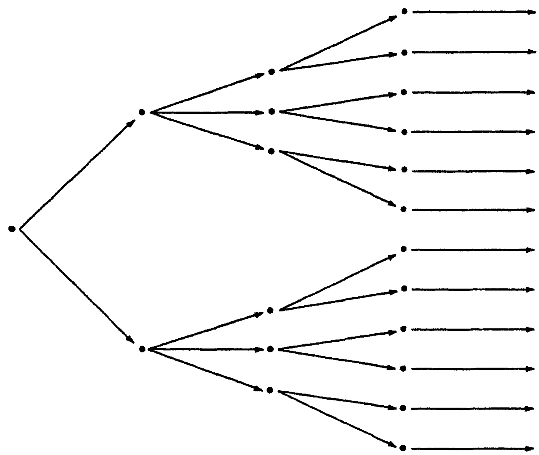


Figure 4-2. Tree diagram listing all round trips between San Francisco and Chicago passing through Salt Lake City that are available to the hypothetical family planning a vacation.

As was apparent from the examples presented in Chapter 3, probability computations generally reduce to the counting of outcomes in an appropriate subspace or subset of an interesting sample space or cartesian product set. While the construction of a tree diagram may simplify the counting of outcomes, such diagrams soon become unwieldy for even small sample spaces. For that reason, other and more efficient counting methods are required. In the following sections some fast counting methods will be defined and derived. These procedures are then used to illustrate probability computations of interesting events in finite sample spaces in which the outcomes are equally likely.

4-2 PERMUTATIONS

Consider a set of books, S_1, to be arranged on a bookshelf. Let this set consist of a single book on anthropology. The number of different ways one may count, order, or permute the individual books in this set consisting of one book is 1. Consider a set, S_2, consisting of an anthropology book and a biology book. The number of different ways one has for counting or ordering these books is 2. They are: (anthropology, biology) and (biology, anthropology). Consider another set, S_3, consisting of an anthropology book, a biology book, and a chemistry book. The number of different ways one may count or order these books is 6. To simplify the remaining discussion, let these books be denoted by the following set of letters $\{A,B,C\}$, A for anthropology, B for biology, and C for chemistry. The six ways for counting or ordering this set of three books are: $\{(A,B,C), (B,C,A), (C,B,A), (C,A,B), (B,A,C),$

$(A,C,B)\}$. Consider another set of books, S_4, consisting of an anthropology book, a biology book, a chemistry book, and a drama book. This set can be denoted, S_4: $\{A,B,C,D\}$. The total number of ways in which these objects can be counted or ordered is 24. Suppose that a fifth book on English is added to the set to give S_5: $\{A,B,C,D,E\}$. The number of different ways for counting or ordering these books is 120. These orders or arrangements are sometimes called *permutations*.

It is quite apparent that the numbers of permutations demonstrate an orderly pattern that is related to the number of objects permuted. In particular,

For one object:	$1 = 1$
For two objects:	$2 = (2)(1)$
For three objects:	$6 = (3)(2)(1)$
For four objects:	$24 = (4)(3)(2)(1)$

And finally,

For five objects:	$120 = (5)(4)(3)(2)(1)$

As these figures suggest, the number of ways for counting or ordering six books or six objects that are all different or distinct is $(1)(2)(3)(4)(5)(6) = 720$. Given seven objects, the number of ways for counting or ordering them is $(1)(2)(3)(4)(5)(6)(7) = 5{,}040$. Consider a class with eight children. The number of ways available for listing the students' names on a class roll is $(1)(2)(3)(4)(5)(6)(7)(8) = 40{,}320$. The number of arrangements of the names of 50 students on a class roster is $(50)(49)(48) \cdots (2)(1)$. This represents an extremely long set of multiplications. Since these kinds of multiplications are common to counting points in sample spaces, a standard mathematical notation for numbers like these has been established. The simplifying notation for the total number of arrangements of 50 distinct objects is $50!$, which is read "50 factorial." (The factors of any number are the numbers that, when multiplied, give the number.) By definition, N factorial is given by

$$N! = (N)(N-1)(N-2)(N-3) \cdots (4)(3)(2)(1)$$

Clearly, $N!$ is a simple corollary following from the fundamental principle of counting. The first place can be filled in N ways. After it is filled, the second place can be filled in $(N-1)$ ways. Thus, the first two places can be filled in $N(N-1)$ ways. Proceeding in this fashion, we find

$$N! = N(N-1)(N-2) \cdots (2)(1)$$

Consider the following set of books, $S_5: \{A,B\ C,D,E\}$ and consider selecting three books from the five books and arranging these books on a shelf. On the first drawing, any one of the five books may be chosen for the first position, leaving four. On the second selection, it is possible to choose any one of the four for the second position. According to the fundamental theorem of counting, the total number of ways of choosing and ordering the two books is $(5)(4)$, or 20. After choosing the

first two, there are still three remaining books to choose from. Applying the fundamental theorem of counting to this selection, we find that the total number is (20) (3), or 60. Interestingly, this result can be written strictly as the ratio of two factorials. To see this, note that

$$60 = (5)\,(4)\,(3) = \frac{(5)\,(4)\,(3)\,(2)\,(1)}{(2)\,(1)} = \frac{5!}{2!} = \frac{5!}{(5-3)!}$$

This result illustrates the following theorem:

Theorem 4-1

The total number of different arrangements of R objects out of a set of N objects is equal to

$$T = \frac{N!}{(N-R)!}$$

Proof. The first position can be filled in N ways, the second position in $(N-1)$ ways, the third position in $(N-2)$ ways, . . . , and the Rth position in $(N-R+1)$ ways. Thus by the extended version of the fundamental theorem of counting,

$$T = N(N-1)(N-2)\cdots(N-R+1)$$

Multiplying the denominator and numerator by $(N-R)!$, we have

$$T = \frac{N(N-1)(N-2)\cdots(N-R+1)(N-R)(N-R-1)\cdots(2)(1)}{(N-R)!}$$

$$= \frac{N!}{(N-R)!}$$

This completes the proof.

As an example of the use of this theorem, consider a class of 20 students from which 4 are going to be selected and used in a learning experiment. The total number of orders in which these students can be tested is

$$T = \frac{20!}{(20-4)!} = \frac{20!}{16!} = (20)\,(19)\,(18)\,(17) = 116{,}280$$

This same number could have been determined simply from the fundamental principle of counting. The total number of ways for choosing the first child is 20. Once this child is removed from the class, the total number of ways for choosing the second child is 19. According to the fundamental principle of counting, the total number of ways for choosing the first 2 children is (20)(19). Once the second child is selected, there are 18 children remaining in the class of which 1 may be chosen. Thus, the total number of ways for choosing the first 3 children is (20)(19)(18), and, finally, the total number of ways for ordering the 4 children is (20)(19)(18)(17).

Table 4-1. Permutations of four different objects compared to the permutations of four objects of which two are identical.

All possible arrangements of four different books		*All possible arrangements of four books in which the drama book has been replaced by an anthropology book*	
ARRANGEMENT NO	ARRANGEMENT	ARRANGEMENT NO	ARRANGEMENT
1	*A B C D*	1	*A B C A*
2	*A B D C*	2	*A B A C*
3	*A C B D*	3	*A C B A*
4	*A C D B*	4	*A C A B*
5	*A D B C*	5	*A A B C*
6	*A D C B*	6	*A A C B*
7	*B A C D*	7	*B A C A*
8	*B A D C*	8	*B A A C*
9	*B C D A*	9	*B C A A*
10	*B C A D*	10	*B C A A*
11	*B D A C*	11	*B A A C*
12	*B D C A*	12	*B A C A*
13	*C A B D*	13	*C A B A*
14	*C A D B*	14	*C A A B*
15	*C B A D*	15	*C B A A*
16	*C B D A*	16	*C B A A*
17	*C D A B*	17	*C A A B*
18	*C D B A*	18	*C A B A*
19	*D A B C*	19	*A A B C*
20	*D A C B*	20	*A A C B*
21	*D B A C*	21	*A B A C*
22	*D B C A*	22	*A B C A*
23	*D C A B*	23	*A C A B*
24	*D C B A*	24	*A C B A*

Consider the set of books S_4 consisting of $\{A\ B,C,D\}$. Suppose the drama book is replaced by an anthropology book identical to the original anthropology book. S_4 now consists of $\{A,B,C,A\}$. For emphasis, remember that both anthropology books are identical: they have the same coloring, the same weight, and the same dimensions. If these books were distinguishable, they could be arranged in 24 different ways. These arrangements are summarized in Table 4-1. The arrangements for the set in which the 2 identical anthropology books have been used are listed in the right column of Table 4-1. This second listing can be derived from the first listing simply by replacing each D by an A.

In the second set of arrangements in the table, it should be noticed that arrangement 1 is the same as arrangement 22 and that arrangement 2 is the same as arrangement 21. Careful inspection will show that each different arrangement is counted twice. Therefore, the total number of different arrangements is 12. In terms of factorials, note that

$$12 = (4)(3) = \frac{(4)(3)(2)(1)}{(2)(1)} = \frac{4!}{2!}$$

This is an illustration of the following theorem:

Theorem 4-2

The total number of arrangements of N objects in which I of them are identical is given by

$$T = \frac{N!}{I!}$$

Proof. The number of arrangements of N distinct objects is $N!$ The number of arrangements of I distinct objects is $I!$ But in this case the I objects are not distinct but identical. This means that each different permutation is counted $I!$ times. Thus, the number of truly different permutations is given by $T = N!/I!$.

This completes the proof.

It should be noted that Theorem 4-1 relates to the arrangement of R objects ($R \leqslant N$) selected from the set of N objects, whereas Theorem 4-2 relates to the arrangement of N objects in which it is known that I are identical. The extension of this theorem to sets with K subsets of identical objects is straightforward. For the general case, Theorem 4-3 suffices. It is presented without proof.

Theorem 4-3

Consider N objects of which I_1 are identical, I_2 are identical, ..., I_K are identical. The total number of different arrangements of these N objects is

$$T = \frac{N!}{(I_1!)(I_2!)\cdots(I_K!)}$$

As an example illustrating this theorem, consider a classroom with 25 students. Suppose the number of female freshmen is given by $I_1 = 7$, the number of male freshmen is given by $I_2 = 3$, the number of male sophomores is given by $I_3 = 8$, and the number of female sophomores is given by $I_H = 7$. If names are neglected, the total number of ways these 25 students may be listed on a class roll in which only their sex and grade level are recorded is given by

$$T = \frac{25!}{(7!)(3!)(8!)(7!)}$$
$$= 2{,}524{,}136{,}472{,}000$$

Note that in determining this astronomical number of permutations, each female freshman is treated as though she were like any other female freshman, with a corresponding statement holding for the remaining three subsets.

According to Theorem 4-3, the number of permutations of N objects in which I are identical and $(N-I)$ are identical is given by

$$T=\frac{N!}{I!(N-I)!}$$

which is sometimes written as.

$$T=\binom{N}{I}$$

The symbol $\binom{N}{I}$ is called a binomial coefficient and can be interpreted as the number of permutations of N objects in which I are identical and $(N-I)$ are identical. In Section 4-3, another interpretation will be given for this symbol.

As an example in which the binomial coefficient may be used, consider a coin that is to be tossed six times in succession. Consider the number of different arrangements of two heads and four tails that are available as possible outcomes for such a sequence of tossings. A little reflection will show that a head on any toss cannot be differentiated from a head on any other toss because a coin cannot change its configuration or shape between tosses. In like manner, every outcome of tails is identical to every other outcome of tails. Using the previous theorem on the total number of arrangements of six objects when four are identical and two are identical, we have

$$T=\binom{6}{2}=\binom{6}{4}=\frac{6!}{2!4!}=\frac{(6)(5)(4!)}{(2)(1)(4!)}=\frac{(6)(5)}{(2)(1)}=15$$

These arrangements are listed in Table 4-2. Without doubt, using the formula is much easier than actually determining and counting each arrangement.

As another example, assume that a coin is to be tossed 100 times. It would be folly to attempt to list all possible outcomes to determine the total number of arrangements of 30 heads and 70 tails. By using the formula we know immediately that the total number of arrangements is equal to

$$T=\binom{100}{30}=\frac{100!}{30!70!}$$

Consider, again, the six consecutive tosses of a coin. The total number of possible outcomes or points in the sample space is equal to (2)(2)(2)(2)(2)(2), or 64. The reason for this is that on the first toss, the coin can land in two different ways, and on the second toss, the coin can also land in two different ways. According to the fundamental theorem of counting, the total number of ways that a coin can fall

on two tosses is (2)(2), or 4. Proceeding in this manner, we can see that the total number of possible outcomes is 2^6, or 64. With this information and the assumption of equally likely outcomes, it is possible to determine the probability of obtaining two heads and four tails if the coin is tossed six times The number of points in the sample space having the property two heads and four tails is 15. The number of points in the sample space is 64. Each of the 15 outcomes is a member of the set $\{X=2\}$. Each outcome can be defined as a permutation of four T's and two H's where each T is identical to the other T's and each H is identical to the other H's Furthermore, the event 4 tails and 2 heads can be characterized as the one event 2 heads. By definition

$$P(2 \text{ heads}) = \frac{n(2H \cap 4T)}{N(S)} = \frac{15}{64}$$

Table 4-2. The 15 possible sequences of 2 heads and 4 tails for the tossing of one coin six times in succession.

Sequence	*Trial number* 1	2	3	4	5	6
1	H	H	T	T	T	T
2	H	T	H	T	T	T
3	H	T	T	H	T	T
4	H	T	T	T	H	T
5	H	T	T	T	T	H
6	T	H	H	T	T	T
7	T	H	T	H	T	T
8	T	H	T	T	H	T
9	T	H	T	T	T	H
10	T	T	H	H	T	T
11	T	T	H	T	H	T
12	T	T	H	T	T	H
13	T	T	T	H	H	T
14	T	T	T	H	T	H
15	T	T	T	T	H	H

4-3 COMBINATIONS

Consider the five books $\{A,B,C,D,E\}$ and the subset $\{A,B,C\}$. This subset represents one *combination* of three books. These three books can be ordered in six different ways. Regardless of the way they are ordered they still represent only one combination of three books. At times one would like to know how many different combinations of three books may be selected from the five books. According to

Theorem 4-1, the total number of arrangements of three books selected from five books is given by

$$T = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{(5)(4)(3)(2)(1)}{(2)(1)} = 60$$

Each combination of books is counted $3! = (3)(2)(1) = 6$ times; therefore, the total number of different combinations of three books out of five books is given by

$$T = \frac{60}{6} = 10 = \frac{5!/(5-3)!}{3!} = \frac{5!}{3!(5-3)!} = \binom{5}{3}$$

Note that $5!/3!(5-3)!$ can be simply denoted in terms of the binomial coefficient $\binom{5}{3}$. In this case, $\binom{5}{3}$ represents the number of different combinations of three objects that can be generated from five objects. In Section 4-2, $\binom{5}{3}$ was also shown to equal the number of permutations of five objects in which three were identical and the remaining two were identical. The meaning of $\binom{5}{3}$ will usually be clear by context. By itself, it is ambiguous.

Theorem 4-4

The total number of different combinations of R objects out of a set of N objects is given by

$$T = \binom{N}{R} = \frac{N!}{R!(N-R)!}$$

Proof. Following the principle outlined in the previous example, we have (the number of permutations of R objects out of N objects) equals (the number of combinations of R objects) multiplied by (the number of times each combination can be permuted) Thus

$$\frac{N!}{(N-R)!} = \binom{N}{R} R!$$

and therefore,

$$\binom{N}{R} = \frac{N!}{R!(N-R)!}$$

This completes the proof.

The number of combinations of N objects out of N objects is equal to 1 since there is only one combination of N objects. By the formula,

$$T = \binom{N}{N} = \frac{N!}{N!(N-N)!} = \frac{N!}{N!0!} = \frac{1}{0!} = 1$$

Furthermore, the number of combinations of 0 objects out of N objects is also equal to 1 and by the formula is given by

$$T = \binom{N}{0} = \frac{N!}{0!(N-0)!} = \frac{N!}{0!N!} = \frac{1}{0!} = 1$$

The only way that these two results can equal 1 and make any sense at all is if 0! equals 1. So that these results do hold and make sense, 0! has, by convention, been defined to equal one. Therefore,

$$\binom{N}{0} = 1 \qquad \text{and} \qquad \binom{N}{N} = 1$$

If a class consists of 10 students, the number of different combinations of 3 students that may serve in a group discussion is given by

$$T = \binom{10}{3} = \frac{10!}{3!7!} = \frac{(10)(9)(8)(7!)}{(3)(2)(1)(7!)} = \frac{(10)(9)(8)}{(3)(2)(1)} = 120$$

Note that the 7! cancels. These kinds of cancellations are very common in computing the values of binomial coefficients. As a result, one can immediately write down the factors to be multiplied and divided by simply starting in the numerator with N and writing in descending order factors equal in number to R. The denominator is given by $R!$. Thus

$$T = \binom{500}{4} = \frac{(500)(499)(498)(497)}{(4)(3)(2)(1)} = 2{,}573{,}031{,}125$$

Also, use can be made of the following theorem to simplify computation:

Theorem 4-5

$$T = \binom{N}{R} = \binom{N}{N-R}$$

Proof.

$$T = \binom{N}{R} = \frac{N!}{R!(N-R)!} = \frac{N!}{(N-R)!R!} = \binom{N}{N-R}$$

This completes the proof.

Thus

$$T = \binom{45}{39} = \binom{45}{45-39} = \binom{45}{6} = \frac{(45)(44)(43)(42)(41)(40)}{(6)(5)(4)(3)(2)(1)}$$
$$= 8{,}145{,}060$$

As another example, consider a class with 30 students, 29 of average ability and 1 who is extremely bright. Suppose an experiment is to be conducted for which 10 students are needed and for which performance in the experiment is related to intelligence. For such an experiment one might hope that this bright student would not be included, provided that the selection is on a random basis. If selection is made at random, what is the probability that this bright student is included in the experiment? According to the definition of probability,

$$P = \frac{n(\text{samples in which bright child is a member})}{N(\text{all samples of size 10})} = \frac{n(A)}{N(S)}$$

The total number of different samples of size 10 that may be chosen from the class of 30 is given by

$$N(S) = \binom{30}{10}$$

Consider now only those samples in which this bright student appears. This means that in these samples there are 9 other children. The total number of ways for selecting 9 children from 29 children is given by $\binom{29}{9}$. Finally, the total number of ways for combining these 9 children with the 1 bright student is given by

$$n(A) = (1)\binom{29}{9}$$

Therefore

$$\begin{aligned} P &= \frac{\binom{29}{9}}{\binom{30}{10}} = \frac{29!/9!20!}{30!/10!20!} \\ &= \left(\frac{29!}{9!20!}\right)\left(\frac{10!20!}{30!}\right) \\ &= \frac{29!\cdot 10\cdot 9!20!}{9!20!\cdot 30\cdot 29!} \\ &= \tfrac{1}{3} \end{aligned}$$

Whereas the evaluation of a binomial coefficient is straightforward, the computations are tedious and create opportunities for error. For these reasons, tables of binomial coefficients have been prepared. One such table is Table A-1 of the Appendix. From this table one can read the value of $\binom{N}{x}$ directly, provided that $N \leqslant 20$ and $x \leqslant 10$. Thus, for example, it is seen that $\binom{15}{3} = 455$. If $x > 10$ then

one can use Theorem 4-5 to determine $\binom{N}{x}$. Thus,

$$\binom{15}{12} = \binom{15}{3} = 455$$

4-4 THE ALGEBRA OF THE SUMMATION SYMBOL $\sum$

In addition to the employment of fast counting methods, statistical procedures rely in considerable measure upon the addition of large masses of data. For example, if X represents the score on a reading test and if 5 students take the test and receive the following set of scores, $\{X_1 = 3,\ X_2 = 6,\ X_3 = 7,\ X_4 = 5,\ X_5 = 10\}$, then the total score for the set is given by $T = X_1 + X_2 + X_3 + X_4 + X_5 = 3 + 6 + 7 + 5 + 10 = 31$. If 37 students take the test, the total score for their set of scores is given by $T = X_1 + X_2 + X_3 + \cdots + X_{35} + X_{36} + X_{37}$. Since sums of this form are common in statistical formulas, it is convenient to have a shorthand symbol to indicate such sums. The commonly employed notation utilizes the Greek letter $\sum$ (sigma) in the following manner:

$$T = X_1 + X_2 + \cdots + X_{36} + X_{37} = \sum_{i=1}^{37} X_i$$

and is read as the sum of X_i from $i = 1$ to 37. For the 5 students,

$$T = \sum_{i=1}^{5} X_i = 3 + 6 + 7 + 5 + 10 = 31$$

The employment of the $\sum$ symbol is quite flexible. For example, for the 5 students, one might like to compute

$$\begin{aligned} T = \sum_{i=1}^{5} X_i^2 &= X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2 \\ &= 3^2 + 6^2 + 7^2 + 5^2 + 10^2 \\ &= 9 + 36 + 49 + 25 + 100 = 219 \end{aligned}$$

Some properties of $\sum$ that find frequent applications in statistical computations are:

1. $$\begin{aligned} \sum_{i=1}^{N} a &= a + a + \cdots + a \\ &= Na \end{aligned}$$

2. $$\begin{aligned} \sum_{i=1}^{N} bX_i &= bX_1 + bX_2 + \cdots + bX_N = b[X_1 + X_2 + \cdots + X_N] \\ &= b \sum_{i=1}^{N} X_i \end{aligned}$$

3. $$\sum_{i=1}^{N} (bX_i + cY_i) = (bX_1 + cY_1) + (bX_2 + cY_2) + \cdots + (bX_N + cY_N)$$
$$= (bX_1 + bX_2 + \cdots + bX_N) + (cY_1 + cY_2 + \cdots + cY_N)$$
$$= b \sum_{i=1}^{N} X_i + c \sum_{i=1}^{N} Y_i$$

4. $$\sum_{i=1}^{N} (X_i - Y_i) = (+1) \sum_{i=1}^{N} X_i + (-1) \sum_{i=1}^{N} Y_i$$
$$= \sum_{i=1}^{N} X_i - \sum_{i=1}^{N} Y_i$$

5. $$\sum_{i=1}^{N} (bX_i - a) = \sum_{i=1}^{N} bX_i + \sum_{i=1}^{N} (-a)$$
$$= b \sum_{i=1}^{N} X_i - Na$$

6. $$\sum_{i=1}^{N} X_i Y_i = X_1 Y_1 + X_2 Y_2 + \cdots + X_N Y_N$$

7. $$\sum_{i=1}^{N} bX_i^2 = bX_1^2 + bX_2^2 + \cdots + bX_N^2 = b[X_1^2 + X_2^2 + \cdots + X_N^2]$$
$$= b \sum_{i=1}^{N} X_i^2$$

8. $$\sum_{i=1}^{N} (bX_i^2 + cY_i^2) = (bX_1^2 + cY_1^2) + (bX_2^2 + cY_2^2) + \cdots + (bX_N^2 + cY_N^2)$$
$$= (bX_1^2 + bX_2^2 + \cdots + bX_N^2) + (cY_1^2 + cY_2^2 + \cdots + cY_N^2)$$
$$= b \sum_{i=1}^{N} X_i^2 + c \sum_{i=1}^{N} Y_i^2$$

9. $$\sum_{i=1}^{N} (X_i - a)^2 = \sum_{i=1}^{N} (X_i^2 - 2aX_i + a^2)$$
$$= \sum_{i=1}^{N} X_i^2 + \sum_{i=1}^{N}(-2aX_i) + \sum_{i=1}^{N} a^2$$
$$= \sum_{i=1}^{N} X_i^2 - 2a \sum_{i=1}^{N} X_i + Na^2$$

10. $$\sum_{i=1}^{N} (X_i - Y_i)^2 = \sum_{i=1}^{N} (X_i^2 - 2X_i Y_i + Y_i^2)$$

$$= \sum_{i=1}^{N} X_i^2 + \sum_{i=1}^{N} (-2X_i Y_i) + \sum_{i=1}^{N} Y_i^2$$

$$= \sum_{i=1}^{N} X_i^2 - 2\sum_{i=1}^{N} X_i Y_i + \sum_{i=1}^{N} Y_i^2$$

Other examples will appear as needed.

4-5 SUMMARY

Probability calculations frequently reduce to the counting of outcomes in a sample space. One of the simplest rules of counting is summarized in the fundamental principle of counting, which states that if the event A_1 can occur in $n(A_1)$ ways, the event A_2 in $n(A_2)$ ways, ..., and the event A_K in $n(A_K)$ ways, then the event $A_1 \otimes A_2 \otimes \cdots \otimes A_K$ can occur in $n(A_1)n(A_2) \ldots n(A_K)$ ways.

A somewhat important special case of the fundamental principle of counting is embodied in the number of permutations or arrangements of N distinct objects and is given by

$$N! = N(N-1)(N-2) \cdots (3)(2)(1)$$

As a special case of this result, it was noted that the total number of different arrangements of R objects out of a set of N objects is given by

$$T = \frac{N!}{(N-R)!} = N(N-1) \cdots (N-R+1)$$

As another special case, it was also noted that the total number of arrangements of N objects in which I of them are identical is given by

$$T = \frac{N!}{I!} = \frac{N(N-1)(N-2) \cdots (3)(2)(1)}{I(I-1)(I-2) \cdots (3)(2)(1)}$$

In the general case in which I_1 objects are identical, I_2 are identical, ..., I_K are identical the number of permutations is given by

$$T = \frac{N!}{(I_1!)(I_2!) \cdots (I_K!)}$$

As a special case of this last result, it was noted that if all of the objects were members of two complementary classes, then the total number of permutations is given by

$$T = \frac{N!}{(I!)(N-I!)} = \binom{N}{I}$$

where $\binom{N}{I}$ is called the binomial coefficient of N upon I. In the context of this discussion, $\binom{N}{I}$ represents the number of permutations of N objects in which I are identical and $(N-I)$ are identical.

Another interpretation that may be given to the binomial coefficient $\binom{N}{R}$ is one that represents the number of combinations of R objects selected from a larger pool of N objects. This interpretation is of considerable value in discussing a special kind of statistical sampling called simple random sampling. This sampling procedure is discussed in considerable detail in Chapters 8 and 9.

Finally, the important addition operator, $\sum$, was introduced. For the set of numbers $\{X_1, X_2, \ldots, X_N\}$, their total is defined as

$$T = \sum_{i=1}^{N} X_i = X_1 + X_2 + \cdots + X_N$$

EXERCISES

***4-1.** In a simple learning experiment, subjects are to be tested until they reach criterion, i.e., until they satisfy the demands of the experimenter. Since the cost of experimentation is high, it was decided to stop the testing after five completed trials Let S represent success in learning the task and F represent failure in learning the task, and draw the tree diagram that represents the complete set of outcomes possible for the study. To help start the tree, note:

$$\begin{tikzcd}
\text{Trial 1} & \text{Trial 2} \\
S & \\
F \arrow[r] \arrow[dr] & S \\
 & F
\end{tikzcd}$$

How many branches does this tree contain? List them.

***4-2.** Repeat Exercise 4-1 if learning is defined to represent two adjacent successes. How many branches does this tree contain? If you can, list the outcomes for the first five trials.

***4-3.** If $P(S) = \frac{1}{3}$ in Exercise 4-1, find the probabilities of the various branches. If the number of trials to success is denoted by X, then X takes on the values $\{0,1,2,3,4,5\}$. Find the probabilities of these outcomes.

***4-4.** A learning experiment is being planned in which subjects are going to be given 10 problems to solve. So as not to permit special preference for any particular problem, the order of presentation is to be different for each subject. How many different orders are possible?

***4-5.** If, in Exercise 4-4, 3 problems tap the same skill while the remaining 7 tap a different skill, how many orderings of skills are there for these 10 problems?

***4-6.** A true-false test is to contain 15 items in which 6 are false and 9 are true. How many different arrangements of true and false answers are possible for this test?

***4-7.** A classroom contains 5 bright students and 20 dull students An experiment is to be conducted for which 10 subjects are needed. If Y represents the number of bright students in the study, find the probability that $Y = 0, 1, 2, 3, 4, 5$, if students are selected at random.

***4-8.** In the experiment of Exercise 4-7, five students are to be assigned to a control condition while the remaining five will be assigned to an experimental condition If X represents the number of bright students in the control condition, find the probability that X takes on the values $\{0,1,2,3,4,5\}$ if students are selected at random.

***4-9.** In a large urban school, the distribution of color blindness by sex of the students is as shown in the following table:

	Males	*Females*	*Total*
Color-blind	20	10	30
Not color-blind	80	140	220
Total	100	150	250

If six students are to be selected for a study in which color discrimination is important, what is the probability that:

(*a*) Two are color-blind?
(*b*) Two are color-blind males?
(*c*) One is a color-blind male and one is a color-blind female?
(*d*) None are color-blind?
(*e*) If three students are assigned to an experimental condition and three are assigned to a control condition, what is the probability that each condition has one color-blind student?

***4-10.** If X represents a score on a reading test while Y represents a score on an arithmetic test, compute the 10 sums illustrated in Section 4-4 for $a = 2$, $b = \frac{1}{5}$, and $c = \frac{1}{5}$ for the following pairs of data.

Pair	*Value of X*	*Value of Y*
1	17	17
2	8	12
3	13	20
4	14	16
5	6	3

5 BERNOULLI VARIABLES

. . . Every time the coin is tossed, each side has an equal chance with the other of winning. If heads wins it is just as likely to win the next time and the next and so on to the thousandth, so that on reasonable grounds a thousand heads in succession are possible, or a thousand tails; for the fact that head wins at any toss does not raise the faintest reasonable probability that tails will win next time. Yet the facts defy this reasoning. Anyone who possesses a halfpenny and cares to toss it a hundred times may find the same side turning up several times in succession; but the total result will be fifty-fifty or as near thereto as does not matter. . . .

The Vice of Gambling and the Virtue of Insurance · George Bernard Shaw. By permission from The Society of Authors, for the Bernard Shaw Estate.

5-1 COIN TOSSING AS A BEHAVIORAL RESEARCH MODEL

Many of the variables encountered in behavioral research have properties that mimic the outcomes that one observes in tossing a coin N times and recording the total number of heads observed in the N tossings. On any one trial, the probability of a head might equal .5 or, if the coin is slightly loaded, .5127. On any one trial the coin must fall heads or tails, so that over 30 trials one may observe a total of $\{0,1,2,3,\ldots,30\}$ heads. If $p =$ 5, one expects to observe 15 heads, while if $p = .5127$, one expects to observe more than 15 heads.

In a similar fashion, a research psychologist could give 30 college sophomores a complex problem to solve in which the probability of solving the problem in five minutes is given by $p = .8$. Thus, in the allotted time one would expect 24 students to solve the problem. In this example, solving the problem corresponds to observing a head upon tossing the coin, but $p = .8$ instead of .5.

As another example, a political scientist could be polling a community to determine the proportion of the electorate that is planning to vote for a particular candidate. In this case a toss of the coin corresponds to the polling of a voter. The occurrence of a head corresponds to the polled person's intention of voting for the candidate. The number of people in a sample of voters who say they intend to vote for the candidate corresponds to the number of heads observed when tossing a coin the number of times equal to the sample size used in the survey.

In like manner, one can think of the outcomes in an experiment in which a drug is being tested to calm manic-depressive patients. For this experiment, the appearance of a head upon tossing a coin corresponds to the efficacy of the drug in calming mental patients. In this case it would be quite unusual to discover that the probability of success is equal to only .5. The researcher in this study would hope that the probability of success would be quite high and close to .9, or even higher. In any case, if the drug is tested on 50 patients, then the number of calmed patients corresponds to the number of heads that appear when an unbalanced coin is tossed 50 times.

All these examples have common elements. In all cases an "all-or-nothing" variable is observed: in coin tossing, the coin must come up heads or tails on each toss; the college sophomores must solve, or fail to solve, the problem; the person polled must state that he plans to vote for, or not to vote for, the candidate; finally, the drug must calm the patient, or fail to calm him. In all these examples, the proportion of individuals in the universe who possess the trait under study corresponds to the probability of a head in a coin-tossing experiment. In this sense, it appears that knowledge concerning coin tossing should have some transfer potential to problem solving, voting preferences, or even drug effectiveness. With this in mind, a rather extensive study of coin tossing will be made. Throughout this discussion and development, heads can be substituted for success in problem solving, voter preference, or even the success that a drug has in calming mental patients. This substitution will not invalidate the discussion. The only reason for using a coin-tossing model is that it simplifies the discussion.

5-2 INTRODUCTION TO BERNOULLI AND BINOMIAL VARIABLES

Up to now, the discussion on probability has centered on the tossing of fair coins. After the following presentation, it will make sense to speak about tossing unfair coins, or coins for which the probability of a head is not $\frac{1}{2}$. In most applications of probability theory to behavioral research, events of interest are not equally likely. For the general development, consider an event A and its complement $\bar{A}$. Let the probability of A be $P(A) = p$ and the probability of $\bar{A}$ be the probability $P(\bar{A}) = 1 - P(A) = 1 - p = q$. A variable that can assume only one of two possible values or states on a given trial with probabilities p and $q = 1 - p$ is called a *Bernoulli variable*. On repeated independent trials, Bernoulli variables generate *binomial variables*. This process is illustrated for coin tossing. For this discussion let A correspond to the appearance of a head and let $\bar{A}$ correspond to the appearance of a tail.

1. Consider a single toss of a coin for which the probability of a head equals p. The possible outcomes of the experiment generate a Bernoulli variable. Their probabilities, summarized in Table 5-1, are correct by definition. For the present discussion, the last column of the table can be ignored.

Table 5-1. Probability distribution of X: number of heads in one toss of a coin.

Outcomes	X	$P(X = x)$	*In binomial formula form*
1 head	1	$p = p$	$\binom{1}{1} p^1 q^{1-1}$
0 head	0	$q = q$	$\binom{1}{0} p^0 q^{1-0}$
Total		$p + q = 1$	$\sum_{x=0}^{1} \binom{1}{x} p^x q^{1-x}$

2. Consider tossing the coin twice and assume that the probability of a head equals p. The possible outcomes of the experiment and the associated probabilities are summarized in Table 5-2.

It is easy to verify these formulas; consider the event 2 heads. The only way this event can occur is if both tosses produce a head. The probability of this event is given by $P(X = 2) = P\{(\text{heads on first trial}) \cap (\text{heads on second trial})\}$. Since tosses are independent, $P(X = 2) = P(H_1)P(H_2)$. Furthermore, since $P(H_1) = P(H_2) = p$, it follows that $P(X = 2) = (pp) = p^2$.

Consider the event 1 head. This event can occur in two mutually exclusive ways. Either the first toss comes up heads and the second tails, or else the first toss comes

Table 5-2. Probability distribution of X: number of heads in two tosses of a coin.

Outcomes	X	$P(X=x)$	*In binomial formula form*
2 heads	2	$pp=p^2$	$\binom{2}{2}p^2q^{2-2}$
1 head	1	$pq+qp=2pq$	$\binom{2}{1}p^1q^{2-1}$
0 heads	0	$qq=q^2$	$\binom{2}{0}p^0q^{2-0}$
Total		$(p+q)^2=1$	$\sum_{x=0}^{2}\binom{2}{x}p^xq^{2-x}$

up tails and the second heads. Therefore, $P(X=1)=P\{$(heads on first trial) $\cap$ (tails on second trial) or (tails on first trial) $\cap$ (heads on second trial)$\}=P(H_1\cap T_2)+P(T_1\cap H_2)$. Since the tosses or trials are independent, it follows that

$$P(X=1)=P(H_1)P(T_2)+P(T_1)P(H_2)$$
$$=(pq)+(qp)=2pq$$

Consider the event 0 heads. The only way that this event can happen is that both tosses come up tails. Since the tosses are independent, it follows that

$$P(X=0)=P\{(\text{tails on first trial})\cap(\text{tails on second trial})\}$$
$$=P(T_1)\,P(T_2)=(qq)=q^2$$

To obtain these results it has been necessary to assume that outcomes are independent and that $P(H)=p$ is the same for all trials.

In deriving the final results, it was necessary to consider all possible ways that the event of interest could happen. For example, to determine $P(X=1)$ it was noted that two different possible outcomes produce one head. Since these two different outcomes are mutually exclusive, their probabilities simply add. This identical argument holds for three, four, five, . . . , or more tosses. Because the argument has been presented for $N=2$, it will not be presented again; however, it must be recalled that what follows depends upon this argument. Since the events $X=2$, 1, 0 exhaust all the possibilities, the probabilities should add to 1. This is easy to prove. From elementary algebra, it is known that

$$p^2+2pq+q^2=(p+q)^2=1^2=1$$

This is recognized as the square of a binomial (a binomial is the sum of two algebraic quantities). In this example, the algebraic quantities are p and q. Probabilities that

satisfy this relationship are referred to as binomial probabilities and variables that possess these probabilities of occurrence are called binomial variables. In fact, the probabilities in Table 5-2 represent individual terms of a binomial raised to a second power. The probabilities in Table 5-1 are simply the terms of the binomial $(p+q)$ raised to the first power.

3. Consider tossing the coin three times. The outcomes and probabilities of the experiment are as summarized in Table 5-3.

Table 5-3. Probability distribution of X: number of heads in three tosses of a coin.

Outcomes	X	$P(X=x)$	*In binomial formula form*
3 heads	3	$ppp = p^3$	$\binom{3}{3} p^3 q^{3-3}$
2 heads	2	$ppq + pqp + qpp = 3p^2q$	$\binom{3}{2} p^2 q^{3-2}$
1 head	1	$pqq + qpq + qqp = 3pq^2$	$\binom{3}{1} p^1 q^{3-1}$
0 heads	0	$qqq = q^3$	$\binom{3}{0} p^0 q^{3-0}$
Total		$(p+q)^3 = 1$	$\sum_{x=0}^{3} \binom{3}{x} p^x q^{3-x}$

Again the probabilities add to 1. The proof is like that for two trials. From elementary algebra it follows that

$$\begin{aligned}
p^3 + 3p^2q + 3pq^2 + q^3 &= (p^3+q^3) + (3p^2q + 3pq^2) \\
&= (p+q)(p^2 - pq + q^2) + 3pq(p+q) \\
&= (p+q)[(p^2 - pq + q^2) + 3pq] \\
&= (p+q)(p^2 + 2pq + q^2) \\
&= (p+q)(p+q)^2 \\
&= (p+q)^3 \\
&= (1)^3 \\
&= 1
\end{aligned}$$

Thus, the probabilities of Table 5-3 represent individual terms of a binomial raised to the third power.

4. Consider tossing the coin four times. Generalizing from the previous examples, one can summarize the outcomes and probabilities of the experiment as shown in Table 5-4.

Table 5-4. Probability distribution of X: number of heads in four tosses of a coin.

Outcomes	$P(X=x)$	*In binomial formula form*
4 heads	p^4	$\binom{4}{4} p^4 q^{4-4}$
3 heads	$4p^3q$	$\binom{4}{3} p^3 q^{4-3}$
2 heads	$6p^2q^2$	$\binom{4}{2} p^2 q^{4-2}$
1 head	$4pq^3$	$\binom{4}{1} p^1 q^{4-1}$
0 heads	q^4	$\binom{4}{0} p^0 q^{4-0}$
Total	$(p+q)^4$	$\sum_{x=0}^{4} \binom{4}{x} p^x q^{4-x}$

As an example, suppose the probability of a head is not equal to $\frac{1}{2}$, but, instead, is equal to $\frac{1}{3}$, and suppose the coin is tossed four times. The probabilities of the various outcomes are as shown in Table 5-5.

Table 5-5. Probability distribution for X: number of heads when $p=\frac{1}{3}$ and $N=4$.

Outcomes	$P(X=x)$	*In binomial formula form*
4	.0123	$(\frac{1}{3})^4$
3	.0988	$4(\frac{1}{3})^3(\frac{2}{3})^1$
2	.2963	$6(\frac{1}{3})^2(\frac{2}{3})^2$
1	.3951	$4(\frac{1}{3})^1(\frac{2}{3})^3$
0	.1975	$(\frac{2}{3})^4$
Total	1.0000	$(\frac{1}{3}+\frac{2}{3})^4=1$

If the coin were fair, the probabilities of the various outcomes would be as shown in Table 5-6.

Table 5-6. Probability distribution for X: number of heads when $p = \frac{1}{2}$ and $N = 4$.

Outcomes	$P(X = x)$	*In binomial formula form*
4	.0625	$(\frac{1}{2})^4$
3	2500	$4(\frac{1}{2})^3(\frac{1}{2})^1$
2	3750	$6(\frac{1}{2})^2(\frac{1}{2})^2$
1	.2500	$4(\frac{1}{2})^1(\frac{1}{2})^3$
0	.0625	$(\frac{1}{2})^4$
Total	1 0000	$(\frac{1}{2} + \frac{1}{2})^4 = 1$

It is convenient at times to make graphical representation of *probability distributions*. Such representations make it easy to see the differences between the probabilities of various outcomes under different conditions. The two probability distributions for $N = 4$ Bernoulli trials with probabilities of success equal to $\frac{1}{3}$ or $\frac{1}{2}$ are shown in Figure 5-1.

5-3 ERRORS IN HYPOTHESIS TESTING

Consider the two probability distributions of Figure 5-1. These probability distributions graphically portray the probabilities of the outcomes for a coin for which $p = \frac{1}{3}$ or $p = \frac{1}{2}$ and for which $N = 4$. Suppose a coin is to be tossed four times and suppose it is known that either $p = \frac{1}{2}$ or $p = \frac{1}{3}$, where p is the probability of a head. Without loss of generality, one can test the assumption that the coin is fair or that $p = \frac{1}{2}$ against the alternative assumption that $p = \frac{1}{3}$. If the hypothesis $p = \frac{1}{2}$ is false, it makes sense to reject it as a true statement; however, if the hypothesis $p = \frac{1}{2}$ is true, it would not be very wise to deny its truth and thereby reject it as a true statement.

Suppose the coin is tossed four times and it is seen that the number of observed heads equals 0 Examination of the probability distribution (Figure 5-1) for the two hypotheses shows that the relative probability of zero heads is greater if $p = \frac{1}{3}$ since $P(X=0|p=\frac{1}{3}) = .1975$, whereas $P(X=0|p=\frac{1}{2}) = .0625$. Because of this, one might argue that the coin is unfair and that p is indeed $\frac{1}{3}$. However, the argument may not be very convincing since it is possible to obtain zero heads with a fair coin. As noted, the probability of this happening by chance alone is .0625. If, indeed, the coin is fair and one concludes that it is not, an error has been made. The probability of making this error is 0625. This is called an error of the first kind, that is,

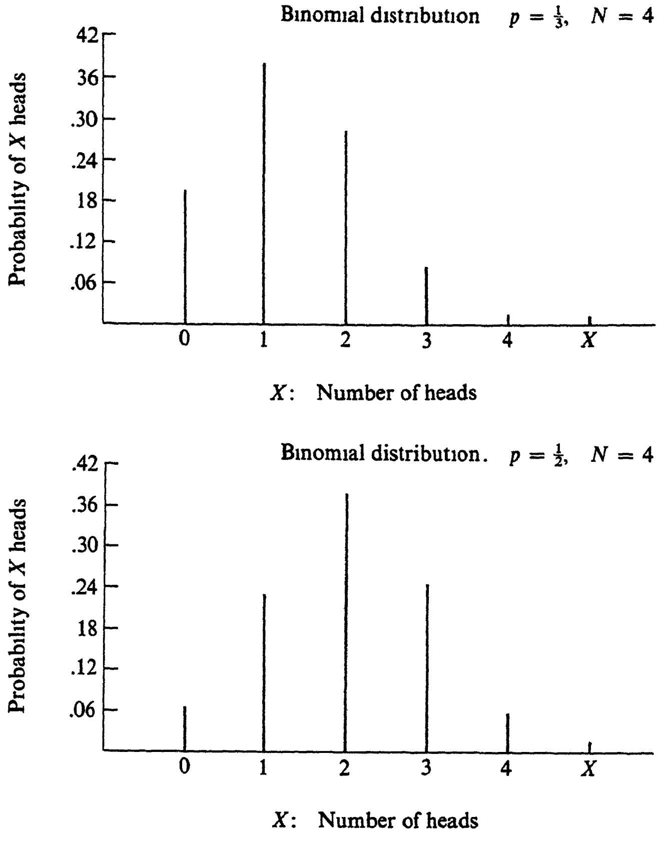


Figure 5-1. Graphic representation of the probability distributions of Tables 5-5 and 5-6.

a *type I error*. It is the denial of a true hypothesis. However, if $p = \frac{1}{3}$ and one concludes that $p = \frac{1}{2}$, a different error is made. This error is called an error of the second kind, that is, a *type II error*. The probability of making this error is $P(X \neq 0 | p = \frac{1}{3}) = 1 - P(X = 0 | p = \frac{1}{3}) = 1 - .1975 = .8025$, a rather high risk of error.

This example illustrates a common problem in statistics of testing a statement or hypothesis about nature against an alternative statement or hypothesis. Since it is physically impossible to collect all the information concerning the truth or falsity of a hypothesis, one must accept the risks entailed in either rejecting a true hypothesis or not rejecting a false hypothesis. In the remainder of the book the groundwork for the solution of this type of problem will be presented. One final goal will be to determine decision rules that minimize the risk of incorrectly rejecting true hypotheses or of not rejecting false hypotheses.

5-4 THE BINOMIAL FORMULA

Reconsider the coin-tossing problem for $N=4$ and $p=\frac{1}{2}$. It was found that the probability of obtaining two heads when a coin is tossed four times is given by

$$P(X=2)=6(\tfrac{1}{2})^2(\tfrac{1}{2})^2=\tfrac{6}{16}=.3750$$

Since $p=\frac{1}{2}$, this probability could also be determined by means of the definition of probability for equally likely events. By definition,

$$P(X=2)=\frac{n(X=2)}{N(S)}$$

The number of points in the sample space is given by $N(S)=(2)(2)(2)(2)=16$. Since the heads as well as the tails are identical, the number of points in the sample space for which there are two heads and two tails is given by the binomial coefficient

$$\binom{4}{2}=\frac{4!}{2!2!}=6$$

Thus

$$P(X=2)=\frac{\binom{4}{2}}{16}=\frac{6}{16}=.3750$$

As can be seen, these two different methods for determining the value of $P(X=2)$ lead to the same result. However, this only occurs because $p=\frac{1}{2}$. If p does not equal $\frac{1}{2}$, the second method is of little value since it cannot be used. Thus, the need for the binomial formula is real.

The binomial formula provides a convenient way for determining the probabilities of events that have the following properties:

1. Each trial, or each experiment, is identical.
2. The probability of the event A (the success event) remains constant and never changes over all trials.
3. Each trial is independent of all others.
4. Events are dichotomous.

The individual trials that satisfy these four conditions are referred to as Bernoulli trials. When speaking of Bernoulli trials, it is customary to refer to the outcome of interest as a *success* and the complement of the outcome of interest as a *failure*. For example, in a guessing game, a correct answer would be referred to as a success and a wrong answer as a failure. In a biological experiment in which drugs are given to guinea pigs, a success would refer to the curing of a disease or the survival of the guinea pig. In a coin-tossing experiment, a head would be referred to as a success and a tail as a failure. The probabilities of events associated with N Bernoulli trials are easily computed by means of the following theorem.

Theorem 5-1

If p equals the probability of a success and $q = (1 - p)$ is the probability of a failure, then the probability that there are x successes in N independent trials is given by

$$P(X=x) = \binom{N}{x} p^x q^{N-x}$$

Proof. In N trials, let the number of successes be denoted by $X = x$ and the number of failures be denoted by $Y = N - x$. For any *one* sequence of independent Bernoulli trials having this structure, the probability is given by

$$[p_1 p_2 \cdots p_x][q_1 q_2 \cdots q_y]$$

Since $p_1 = p_2 = \cdots = p_x = p$ and $q_1 = q_2 = \cdots = q_y = q$, this probability reduces to

$$p^x q^y = p^x q^{N-x}$$

As stated, this represents the probability of only one possible sequence of x successes and $N - x$ failures. The total number of different sequences is equal to the total number of different arrangements of x identical objects and $(N - x)$ different but identical objects. The total number of such permutations is given by

$$T = \frac{N!}{x!(N-x)!} = \binom{N}{x}$$

Thus, the probability of interest is given by

$$P(X=x) = \binom{N}{x} p^x q^{N-x}$$

This completes the proof.

As an example of the use of this formula, consider the following: suppose 20 percent of the students who take a beginning course in statistics fail; what is the probability that at most 2 students in a class of 15 fail? The event, at most 2, is equivalent to the union of the following events: {0 failing, 1 failing, or 2 failing}. Therefore, the probability of at most 2 is the same as

$$P(\text{at most 2 failing}) = P(X \leqslant 2) = P[(X=0) \cup (X=1) \cup (X=2)]$$

Since these events are mutually exclusive,

$$P(X \leqslant 2) = P(X=0) + P(X=1) + P(X=2)$$

At this point it must be assumed that the individual performances in the class are independent, although they might not be. For example, one bright student might be giving assistance to a less bright student and therefore would help this student along. Even if the events are not strictly independent, the approximation to the binomial may be close enough. If it is true that the events are independent, then

$$P(X=2) = \binom{15}{2}(.2)^2(.8)^{13} = .2309$$

$$P(X=1) = \binom{15}{1}(.2)^1(.8)^{14} = .1319$$

$$P(X=0) = \binom{15}{0}(.2)^0(.8)^{15} = .0352$$

so that the probability of interest is given by

$$P(X \leqslant 2) = .0352 + .1319 + .2309 = .3980$$

5-5 HYPOTHESIS TESTING—AN EXAMPLE

Consider a lady who claims that she has extrasensory perception (ESP) and that she can identify the color of a card that an experimenter is looking at while she is in another room. That is, if the experimenter in one room is looking at a black card, she can correctly identify it as black even though she is in another room. If the card is red, she can do the same. Naturally, one would like to know whether she really has this ability or whether it is just imagined. The way to determine this is to conduct a test. The purpose of the test should be either to establish her ability to identify the cards or to prove that she cannot identify the cards. Being somewhat skeptical concerning her ability, most experimenters would insist that she prove her claim, and in order to do this she will have to disprove the experimenter's claim. The experimenter's claim is that she does not have the ability. Let this hypothesis to be tested be denoted by H_0. Thus

H_0: she does not have the ability

Naturally, she is offering an alternative hypothesis. Let this alternative hypothesis be denoted as H_1. Thus

H_1: she has the ability

If she really doesn't have the ability, the probability that she will guess correctly on any one trial is $\frac{1}{2}$, since her only possible choices are correct or incorrect. Since the trials are to be independent, probabilities of correct guesses can be computed by the binomial formula. As a result, the experimental hypothesis that she does not have the ability is equivalent to the statistical hypothesis

$$H_0: p = \tfrac{1}{2}$$

Her hypothesis is equivalent to the following alternative statistical hypothesis:

$$H_1: p > \tfrac{1}{2}$$

Suppose the experimenter decides to give her only two trials. The possible outcomes are: {(correct, correct), (correct, wrong), (wrong, correct), (wrong, wrong)}.

Consider the probabilities associated with all possible outcomes, and in order to do this let X equal the number of correct decisions. Possible values for X are X: {0,1,2}. Probabilities of these outcomes can be computed by means of the binomial formula. Thus

$$P(X=2|p=\tfrac{1}{2}) = \binom{2}{2}(\tfrac{1}{2})^2(\tfrac{1}{2})^0 = \tfrac{1}{4} = .25$$

$$P(X=1|p=\tfrac{1}{2}) = \binom{2}{1}(\tfrac{1}{2})^1(\tfrac{1}{2})^1 = \tfrac{2}{4} = .50$$

$$P(X=0|p=\tfrac{1}{2}) = \binom{2}{0}(\tfrac{1}{2})^0(\tfrac{1}{2})^2 = \tfrac{1}{4} = .25$$

Even if she doesn't have the ability, the probability that she will make two correct decisions is .25; that is, the risk of accepting her claim when she really doesn't have the ability is 1 in 4. That is a rather high risk and one that most experimenters would not be willing to accept. Without doubt, two trials are not sufficient.

Suppose the experimenter gives her four trials. Then the possible values for X are X: {0,1,2,3,4}. With these outcomes the probabilities are given by the following:

$$P(X=4|p=\tfrac{1}{2}) = \binom{4}{4}(\tfrac{1}{2})^4(\tfrac{1}{2})^0 = \tfrac{1}{16} = .0625$$

$$P(X=3|p=\tfrac{1}{2}) = \binom{4}{3}(\tfrac{1}{2})^3(\tfrac{1}{2})^1 = \tfrac{4}{16} = .2500$$

$$P(X=2|p=\tfrac{1}{2}) = \binom{4}{2}(\tfrac{1}{2})^2(\tfrac{1}{2})^2 = \tfrac{6}{16} = .3750$$

$$P(X=1|p=\tfrac{1}{2}) = \binom{4}{1}(\tfrac{1}{2})^1(\tfrac{1}{2})^3 = \tfrac{4}{16} = .2500$$

$$P(X=0|p=\tfrac{1}{2}) = \binom{4}{0}(\tfrac{1}{2})^0(\tfrac{1}{2})^4 = \tfrac{1}{16} = .0625$$

The experimenter might now wish to place a restriction upon the test, and that restriction says her claim will be established only if she gets four correct answers. The probability that she could do this even if she doesn't have ESP is given by $P(X=4|p=\tfrac{1}{2}) = .0625$, or 1 in 16, which might not be as low as most experimenters would like but is the best that can be done in this case. Undoubtedly she will be unhappy with this decision rule. Most likely she will now present the argument that she can't do it all the time but she can do it most of the time. She will insist that the experimenter grant her claim if she makes three or four correct identifications! If the experimenter grants her this concession, then the probability that he could accept her claim when she doesn't really have the ability is given by

$$P(X=4|p=\tfrac{1}{2}) + P(X=3|p=\tfrac{1}{2}) = .0625 + .2500 = .3125$$

which is even worse than the risk he entailed with only two trials. As might be expected, this rule would not be acceptable to him.

Suppose he decides to give her 10 trials. This puts him in a position where he can determine a reasonable decision rule, one that might be acceptable to her. Certainly if she made nine or ten correct identifications, he would not hesitate to grant her claim. But suppose she made eight or seven correct choices; then he might not want to. In order to establish a good test, consider a set of possible decision rules and the probabilities that these rules would entail incorrectly leading the experimenter to accept her claim when she doesn't have the ability. These rules are summarized in Table 5-7.

Table 5-7. Decision rule for the lady who guesses card colors.

	Possible rule	*Probability of concluding lady has* ESP *when she really doesn't*
I.	Accept her claim if $X = 10$	$P(X=10) = \sum_{x=10}^{10} \binom{10}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} = \frac{1}{1024} = 0010$
II.	Accept her claim if $X = 10$ or 9	$P(X \geqslant 9) = \sum_{x=9}^{10} \binom{10}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} = \frac{11}{1024} = .0107$
III	Accept her claim if $X = 10$, 9, or 8	$P(X \geqslant 8) = \sum_{x=8}^{10} \binom{10}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} = \frac{56}{1024} = .0547$
IV	Accept her claim if $X = 10$, 9, 8, or 7	$P(X \geqslant 7) = \sum_{x=7}^{10} \binom{10}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} = \frac{176}{1024} = .1719$

She probably would object to rules I and II, declaring them too stringent. She may continue to insist that she does have the ability but cannot guess correctly all the time and, therefore, rules I and II, which require that she do it correctly a great proportion of the time, are unacceptable to her. She might continue to object to rule III, as this also places a high demand on her ability. Rule IV she might accept, but the experimenter clearly would not. The risk, .1719, of granting her claim when she doesn't have the ability is too high. About the worst he would accept is rule III.

For this rule, the risk of a wrong conclusion is about .05. Unless she will accept this rule, he probably won't test her.

Suppose she does consent to perform the experiment and suppose she identifies seven cards correctly. Because the decision rule says she must identify eight, nine, or ten correctly before he accepts her claim, he must conclude that she does not have the ability to identify the cards even though she did it correctly in seven out of ten trials.

The statistical analysis of the experiment involving ESP is based on a weak but frequently performed statistical test called the binomial test. A special case of this test called the sign test is described below.

5-6 THE SIGN TEST

The sign test is one of the easiest-to-use statistical tests available to the behavioral researcher. This popular test is dependent upon the construction of a variable that has a binomial probability distribution with $p = \frac{1}{2}$, provided that the hypothesis under test is true. To see how this variable is constituted, consider the following experiment.

Twenty-four fifth grade remedial readers were given a standardized reading test and then ranked according to final test scores. Once the students were ranked, the two with the lowest scores were paired in what is commonly referred to as a matched pair This pairing was repeated for the next two lowest scores and continued until 12 pairs of students were created. On the basis of a coin toss, one member of each pair was assigned to an experimental reading program consisting of 4 weeks of special reading training on a typewriter connected to a high-speed computer and programmed with special reading materials. The remaining student of each pair was given normal training in the usual classroom setting. Following the training, all 24 students were given a specially designed reading test. The results of the testing are shown in Table 5-8.

If the special training has no effect, or if the hypothesis is true, then one would expect to find that in half the pairs the experimental subject scored higher than the subject placed in the control condition. Operationally, this would mean that the probability of a positive difference or a plus sign should be given by $P(+) = \frac{1}{2}$. Since each pair of scores is independent of the others, it follows that if X represents the number of + deviations, then the probability distribution of X is given by

$$P(X = x) = \binom{12}{x} (\tfrac{1}{2})^x (\tfrac{1}{2})^{12-x}$$

with X: $\{0,1,2,\ldots, 12\}$.

With this formula, it is easy to set up an appropriate decision rule. If one decided to adopt a decision rule so that the total probability of saying that a difference exists when none does is less than .05, then one should decide to reject the hypothesis of no difference if the number of positive deviations is small in number. By trial-and-error methods, as illustrated in the previous example, it is easy to show that the

Table 5-8. Scores on a reading test for 12 ranked matched pairs of fifth grade remedial reading students.

Pair	*Type of student training*		*Difference*	*Algebraic sign of the difference*
	NORMAL	TYPEWRITER		
1	23	25	−2	−
2	17	22	−5	−
3	30	24	+6	+
4	27	23	+4	+
5	28	26	+2	+
6	30	31	−1	−
7	35	17	+18	+
8	26	28	−2	−
9	31	25	+6	+
10	35	33	+2	+
11	37	39	−2	−
12	36	30	+6	+

appropriate decision rule for this test is: reject the hypothesis if $X \in \{0,1,2\}$, since

$$P(X \leqslant 2) = \binom{12}{0}(\tfrac{1}{2})^0(\tfrac{1}{2})^{12} + \binom{12}{1}(\tfrac{1}{2})^1(\tfrac{1}{2})^{11} + \binom{12}{2}(\tfrac{1}{2})^2(\tfrac{1}{2})^{10}$$

$$= \tfrac{79}{4096} = .019$$

while $P(X \leqslant 3) = \frac{299}{4096} = .07$

In this case the number of positive differences is given by $X = 7$, so the hypothesis is not rejected. Operationally this means that the special training on the typewriter is just as effective as the normal classroom training.

Tables for the sign test have been prepared and are included in the Appendix as Table A-2. This table is entered with N, the sample size, and the *maximum* risk of a type I error that one is willing to tolerate. Possible *maximum* values are {.01, .05, .10, .25}, provided that one tests $H_0: p = \frac{1}{2}$ versus the alternative $H_1: p \neq \frac{1}{2}$. To test $H_0: p = \frac{1}{2}$ versus $H_1: p < \frac{1}{2}$, divide the tabled risk values in half. This same division by $\frac{1}{2}$ is valid for testing $H_0: p = \frac{1}{2}$ versus $H_1: p > \frac{1}{2}$.

For the examples in this chapter, Table A-2 shows that:

1. For the ESP study, $N = 10$, and for the test of $H_0: p = \frac{1}{2}$ versus $H_1: p > \frac{1}{2}$, H_0 should be rejected at a 5 percent risk of a type I error if $X = \{0,1\}$, where X equals the number of incorrect choices. Since $X = 3$, H_0 is not rejected.
2. For the matched pair reading study, $N = 12$, and for the test of $H_0: P(+) = \frac{1}{2}$ versus $H_1: P(+) < \frac{1}{2}$, H_0 should be rejected at a 5 percent risk of a type I error

if $X = \{0,1,2\}$, where X equals the number of positive differences Since $X = 7$, H_0 is not rejected.

5-7 INTRODUCTION TO SAMPLING THEORY—CONTINUATION

As was stated in Chapter 4, sampling without replacement is the usual sampling scheme for behavioral studies. However, the mathematics of the correct sampling procedures are complex and thus simpler procedures are desirable. A good approximation is available through the binomial formula. In this section the approximate binomial procedures will be compared to the correct procedure, which is computed by a formula called the hypergeometric formula. The differences in these procedures will be illustrated by examples.

Sampling with replacement

Consider a classroom with 5 boys and 15 girls from which a sample of 5 children is to be selected. Consider the case that a child is selected and after his sex is determined he is returned to the classroom, where he could be selected again. If X equals the number of boys in the sample, the possible values for X are X: $\{0,1,2,3,4,5\}$. Since the selected subjects are returned to the classroom after each drawing, the probability of selecting a boy over every trial is given by $P(B) = n(B)/N(S) = \frac{5}{20} = \frac{1}{4}$ and the probability of selecting a girl over every trial is given by $P(G) = n(G)/N(S) = \frac{15}{20} = \frac{3}{4}$. In addition, sampling is independent under this condition because what happens on previous selections in no way influences subsequent selections. The probabilities of interest are shown in Table 5-9.

Table 5-9. Probability distribution of X: number of boys in sample for $N = 5$ and $p = \frac{1}{4}$.

Outcome	$P(X=x)$	*Binomial formula*	*Probability*
5	$P(X=5)$	$\binom{5}{5}\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right)^0$	.0100
4	$P(X=4)$	$\binom{5}{4}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^1$	.0146
3	$P(X=3)$	$\binom{5}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2$	.0879
2	$P(X=2)$	$\binom{5}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3$	.2637
1	$P(X=1)$	$\binom{5}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^4$	.3955
0	$P(X=0)$	$\binom{5}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^5$	.2373
Total	$\sum_{x=0}^{5} P(X=x)$	$\sum_{x=0}^{5}\binom{5}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{5-x}$	1.0000

Sampling without replacement

Consider the same problem, but without replacement. Under these conditions, the probability of choosing a boy does not remain constant. For example, if a boy is chosen on the first drawing, the probability of choosing a boy on the second drawing is $\frac{4}{19}$. However, if a girl is chosen on the first drawing, the probability of choosing a boy on the second drawing is $\frac{5}{19}$. In either case, it is clearly not $\frac{5}{20}$. Sampling without replacement must account for this. In order to determine the probabilities associated with the outcomes under this condition, it is necessary to fall back on the equally likely definition of probability.

Since the probabilities of individual events are not equally likely, it is necessary to introduce the concept of equally likely events somewhere into the model. For this development, consider the complete set of different samples of size 5 that may be selected from this population. This number is given by $T=\binom{20}{5}$. If each of *these* samples is equally likely, then the probability of selecting any one sample of size 5 is given by

$$p=\frac{1}{T}=\frac{1}{\binom{20}{5}}$$

Thus, to find the probability of any event of interest it is necessary only to count the number of samples in the entire set of samples that have the property of interest and then to divide this total by T, the total number of possible samples.

Note that the statement "$X=0$" is incomplete. One should really state that "$X=0$ and $Y=5$," where X equals the number of boys and Y equals the number of girls. According to the definition of probability,

$$P\{(X=0)\cap(Y=5)\}=\frac{n\{(X=0)\cap(Y=5)\}}{N(S)}$$

As was already shown,

$$N(S)=\binom{20}{5}$$

To obtain the numerator, consider the separate events. The number of ways to choose zero boys from five boys is $\binom{5}{0}$. After the boys are chosen, the number of ways to choose five girls from 15 girls is $\binom{15}{5}$. If the first event can be done in $\binom{5}{0}$ ways and if the other event can be done in $\binom{15}{5}$ ways, then according to the fundamental principle, the total number of ways for doing both is $\binom{5}{0}\binom{15}{5}$.

Therefore,

$$P(X=0)=\frac{\binom{5}{0}\binom{15}{5}}{\binom{20}{5}}=.1937$$

Using the same reasoning, one can obtain the probabilities for the remaining events. The probability distribution of interest is shown in Table 5-10. Note that the probabilities calculated under replacement differ from those calculated without replacement. But the differences are not very large. Notice that the correct probability

Table 5-10. Probability distribution of X: number of boys in a sample of size 5 for a universe with 5 boys and 15 girls.

Outcome	$P(X=x)$	*Hypergeometric formula*	*Probability*
5	$P(X=5)$	$\frac{\binom{5}{5}\binom{15}{0}}{\binom{20}{5}}$	.0001
4	$P(X=4)$	$\frac{\binom{5}{4}\binom{15}{1}}{\binom{20}{5}}$	.0048
3	$P(X=3)$	$\frac{\binom{5}{3}\binom{15}{2}}{\binom{20}{5}}$	.0677
2	$P(X=2)$	$\frac{\binom{5}{2}\binom{15}{3}}{\binom{20}{5}}$	.2935
1	$P(X=1)$	$\frac{\binom{5}{1}\binom{15}{4}}{\binom{20}{5}}$	.4402
0	$P(X=0)$	$\frac{\binom{5}{0}\binom{15}{5}}{\binom{20}{5}}$	.1937
Total	$\sum_{x=0}^{5} P(X=x)$	$\sum_{x=0}^{5}\frac{\binom{5}{x}\binom{15}{5-x}}{\binom{20}{5}}$	1.0000

theory for experimentation is described by sampling without replacement, but the probabilities are difficult to compute. Since the binomial probabilities are very close to these and are easier to compute, they are used as approximations to the exact probabilities. This approximation is valid only if the ratio of the sample size to the universe size is less than 10 percent. While this is not true in this case, the approximation is still good.

5-8 THE HYPERGEOMETRIC FORMULA

The exact probabilities of the previous section are referred to as hypergeometric probabilities and are described by the hypergeometric formula. If

N_1 = number of A's in the universe

N_2 = number of $\bar{A}$'s in the universe

x_1 = number of A's in the sample

x_2 = number of $\bar{A}$'s in the sample

then

$$P(X_1 = x_1 \cap X_2 = x_2) = \frac{\binom{N_1}{x_1}\binom{N_2}{x_2}}{\binom{N_1 + N_2}{x_1 + x_2}}$$

As another example of the use of this formula, consider a classroom of 14 children: 8 Caucasians and 6 Negroes. Suppose a sample of size 5 is to be drawn from the classroom and used in a learning experiment. Let X equal the number of Negroes in the sample. Possible values for X are X: {0,1,2,3,4,5}. The probabilities of these events can be determined directly from the hypergeometric formula or from the basic definition of probability for equally likely events. The latter method will be used mainly to reinforce the theory behind the hypergeometric formula. As a check, it is suggested that the reader verify these results by substitution into the hypergeometric formula. By definition,

$$P(X = x) = \frac{n(\text{samples with } x \text{ Negroes})}{N(\text{samples of size 5})} = \frac{n(X = x)}{N(S)}$$

Consider the denominator. The total number of different samples of size 5 that may be drawn from the population of 14 students is

$$N(S) = \binom{14}{5}$$

Consider the event $X = 0$. The only way this can occur is if the number of Negroes in the sample is 0 and the number of Caucasians in the sample is 5. The total number of ways for choosing 0 Negroes from 6 Negroes is $\binom{6}{0}$. The total number of ways

for choosing 5 Caucasians from the total of 8 Caucasians is $\binom{8}{5}$ According to the fundamental principle of counting, the total number of ways available for choosing 0 Negroes and 5 Caucasians is given by the product $\binom{6}{0}\binom{8}{5}$. Therefore,

$$P(X=0)=\frac{\binom{6}{0}\binom{8}{5}}{\binom{14}{5}}=\frac{4}{143}=.0280$$

Table 5-11. Probability distribution of X number of Negroes in a sample of size 5 from a class containing six Negroes and eight Caucasians.

Outcome	$P(X=x)$	*Hypergeometric formula*	*Probability*
5	$P(X=5)$	$\frac{\binom{6}{5}\binom{8}{0}}{\binom{14}{5}}$	0030
4	$P(X=4)$	$\frac{\binom{6}{4}\binom{8}{1}}{\binom{14}{5}}$	0599
3	$P(X=3)$	$\frac{\binom{6}{3}\binom{8}{2}}{\binom{14}{5}}$	2797
2	$P(X=2)$	$\frac{\binom{6}{2}\binom{8}{3}}{\binom{14}{5}}$	.4196
1	$P(X=1)$	$\frac{\binom{6}{1}\binom{8}{4}}{\binom{14}{5}}$	.2098
0	$P(X=0)$	$\frac{\binom{6}{0}\binom{8}{5}}{\binom{14}{5}}$	0280
Total	$\sum_{x=0}^{5} P(X=x)$	$\sum_{x=0}^{5}\frac{\binom{6}{x}\binom{8}{5-x}}{\binom{14}{5}}$	1.0000

In like manner, the total number of ways for choosing 1 Negro from 6 Negroes is $\binom{6}{1}$. The total number of ways for choosing 4 Caucasians from 8 Caucasians is given by $\binom{8}{4}$. Therefore, by the fundamental principle of counting, the total number of ways for choosing a sample with 1 Negro and 4 Caucasians is given by $\binom{6}{1}\binom{8}{4}$ and the probability of this event is

$$P(X=1)=\frac{\binom{6}{1}\binom{8}{4}}{\binom{14}{5}}=\frac{30}{143}=.2098$$

The probabilities of interest are as summarized in Table 5-11.

5-9 THE IRWIN-FISHER EXACT TEST

Tables for hypergeometric probabilities have been prepared and are included in the Appendix as Table A-3. The following example shows the usefulness of this table in testing the hypothesis that conditional probabilities in a 2×2 table are identical.

Eight rats were exposed to food deprivation for two weeks and were then tested on maze running. At the same time, five normally fed rats were tested. The results of the experiment in terms of the successful learning of the maze are as shown in Table 5-12. As is suggested by the 2×2 table results, the researcher wishes to know

Table 5-12. Learning performance of normally fed and food-deprived rats on maze task. Success defined in numbers of trials taken to learn the maze.

	Food-deprived rats	*Normally fed rats*	*Total*
Learned maze in 10 trials or less	6	1	7
Learned maze in 11 trials or more	2	4	6
Total	8	5	13

whether food deprivation influences rate of learning. If one hypothesizes that "Maze learning is independent of previous food deprivation condition," then one can use Table A-3 to test this hypothesis. One enters the table with

N = total sample size
S_1 = smallest marginal row or column total
S_2 = next smallest row or column total
X = frequency in the cell defined by S_1 and S_2

For the observed data, $N = 13$, $S_1 = 5$, $S_2 = 6$, and $X = 4$. If the probability associated with these values is less than .05, the hypothesis of independence or no difference in effects is rejected. For these numbers, the total is $p = .103$, so the conclusion is that food deprivation has no effect upon the learning of the maze.

It should be noted that this example represents a formal test of the hypotheses

$$H_0: P(A|B) = P(A|\bar{B})$$

versus

$$H_1: P(A|B) \neq P(A|\bar{B})$$

for which a rule of thumb was presented in Section 3-10. Whenever the total sample size is less than 15, one can use Table A-3 to make a direct test of the hypothesis of independence or no effect. Procedures for $N \geqslant 15$ are presented in Chapters 11, 12, and 16. When direct use is made of the hypergeometric probabilities, this test is referred to, in the literature, as the Irwin–Fisher exact test of significance.

The manner in which Table A-3 is constructed makes it easy to test that one conditional probability is larger than another. While this procedure is not described, the reader should have little difficulty in using Table A-3 for this kind of alternative hypothesis.

5-10 SUMMARY

Many of the variables of behavioral research have properties similar to the variable X that specifies the number of heads that appear when a coin is tossed N times. While the probability of a head on any one trial is $\frac{1}{2}$, this same statement does not generally hold true for the related variables encountered in behavioral research. For example, a sample of movie goers attending a foreign film may be asked, "Do you attend foreign films because they contain messages of social significance more often than American films? Yes______ No______." The responses {yes, no} correspond to the coin-tossing outcomes {heads, tails}. The probability of a "yes" response corresponds to the probability of a "head." Variables that have this "all-or-nothing" property on single trials are called Bernoulli variables.

Over repeated trials, Bernoulli variables generate binomial variables. Thus, if Y is a Bernoulli variable with $P(Y = A) = p$ and $P(Y = \bar{A}) = q$, and if $X = Y_1 + Y_2 + \cdots + Y_N$ where Y_i = outcome on the ith trial, and if the trials are independent, then X is a binomial variable with probability given by

$$P(X = x) = \binom{N}{x} p^x q^{N-x}$$

Binomial variables are encountered in statistical hypothesis-testing models and are used to evaluate the probabilities of making type I and type II errors. In the study concerning the social significance of foreign films, it might be hypothesized that one of the reasons people go to foreign films is that they regard the films as socially important. Operationally, this statement or hypothesis about the reasons why people attend foreign films can be translated into the following statistical hypothesis:

$$H_1: P(\text{yes}) = \tfrac{9}{10}$$

which states essentially that 90 percent of foreign film fanciers consider these films to be socially significant. On the other hand, foreign film advocates might just enjoy foreign films because they are technically and visually interesting, and for these individuals it may really be that

$$H_0: P(\text{yes}) = \tfrac{1}{2}$$

This states essentially that foreign film goers do not particularly view foreign films from a strict social-significance point of view.

Note that either H_0 or H_1 is true. Since it is generally impossible to obtain information from the entire population of foreign film devotees, one must rely on the information obtained from a sample to decide between H_0 and H_1. If, on the basis of sample evidence, it is concluded that H_1 is true when really H_0 is true, a type I error has been committed; if, on the other hand, it is concluded that H_0 is true when really H_1 is true, a type II error has been committed. If these conclusions are based upon the following decision rule to conclude that H_1 is true when $X \geqslant 35$ in a sample of 50 foreign film goers, then

$$\begin{aligned} P(\text{type I error}) &= \binom{50}{35}\left(\frac{1}{2}\right)^{35}\left(\frac{1}{2}\right)^{15} + \binom{50}{36}\left(\frac{1}{2}\right)^{36}\left(\frac{1}{2}\right)^{14} + \cdots + \binom{50}{50}\left(\frac{1}{2}\right)^{50}\left(\frac{1}{2}\right)^{0} \\ &= \sum_{x=35}^{50}\binom{50}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{50-x} \\ &= .002 \end{aligned}$$

and

$$\begin{aligned} P(\text{type II error}) &= \binom{50}{0}\left(\frac{9}{10}\right)^{0}\left(\frac{1}{10}\right)^{50} + \binom{50}{1}\left(\frac{9}{10}\right)^{1}\left(\frac{1}{10}\right)^{49} \\ &\qquad + \cdots + \binom{50}{34}\left(\frac{9}{10}\right)^{34}\left(\frac{1}{10}\right)^{16} \\ &= \sum_{x=0}^{34}\binom{50}{x}\left(\frac{9}{10}\right)^{x}\left(\frac{1}{10}\right)^{50-x} \\ &= .0001 \end{aligned}$$

Binomial probability theory serves as a good *approximation* for sampling from finite populations in which selected objects are not returned for possible repeat selection. The binomial formula is valid when

1. Each trial or each experiment is conducted under identical conditions.
2. The probability of the event A remains constant over all trials.
3. Each trial is independent of all others.
4. Events are dichotomous.

When sampling without replacement from finite populations, the probability of the event A changes with each trial and furthermore, the trials are not independent of one another. The appropriate probability numbers are defined, in this case, by the hypergeometric formula. If

$$\frac{n}{N} = \frac{x_1 + x_2}{N_1 + N_2} < 10 \text{ percent}$$

then the hypergeometric probabilities can be approximated from the binomial formula by using

$$p = \frac{x_1}{N} \qquad \text{and} \qquad q = \frac{x_2}{N}$$

so that

$$P(X_1 = x_1 \cap X_2 = x_2) = \binom{n}{x_1} p^{x_1} q^{x_2} = \binom{n}{x_2} p^{x_1} q^{x_2}$$

EXERCISES

***5-1.** Each item of a multiple choice test consisting of six items has been correctly answered in the past by 80 percent of the tested subjects. If a student is selected at random and is given the test, what is the probability he will get

(*a*) No correct answers?
(*b*) No less than five correct answers?
(*c*) At most four correct answers?
(*d*) At least four correct answers?
(*e*) Six correct answers?

***5-2.** If the test in Exercise 5-1 is given to 120 students, how many students are expected to get

(*a*) No correct answers?
(*b*) No less than five correct answers?
(*c*) At most four correct answers?
(*d*) At least four correct answers?
(*e*) Six correct answers?

***5-3.** Seven maze-bright rats are to be trained to run a T maze. What is the probability distribution of correct runs for these rats on the first trial? A T maze has only one choice point so that a tested rat must make a right- or left-hand turn.

***5-4.** From past experience it is known that by the fourth run of the T maze of Exercise 5-3, the probability distribution of correct runs is binomial with $p = .9$. Compute the probabilities for the number of rats that will run the maze correctly with $N = 7$.

***5-5.** Plot the two distributions of Exercises 5-3 and 5-4. What does the graph suggest about learning?

***5-6.** Using the results of Exercises 5-3 and 5-4, what can you say about

(*a*) $P(X = 7|\text{first run})$ and $P(X = 7|\text{fourth run})$?
(*b*) $P(X = 6|\text{first run})$ and $P(X = 6|\text{fourth run})$?

***5-7.** In a certain community, bond issues are passed only if two-thirds or more voters agree to the passage. In the last vote on the passage of a bond issue to start a junior college, the yes vote was equal to 60 percent. In the meantime, a campaign was carried on to push the bond issue through. Two weeks prior to the election, 10 people were polled on how they would vote on the bond issue. Nine polled citizens said that they were going to vote for passage.

(*a*) Compute $P(X = 9|p_0 = \frac{6}{10})$.
(*b*) Compute $P(X = 9|p_1 = \frac{2}{3})$.
(*c*) What do these two probabilities represent?
(*d*) Do you think the bond issue will pass? Explain Binomial probability computations can be simplified by using logarithms Thus

$$\log P(X = x) = \log \binom{N}{x} + x \log p + (N - x) \log q$$

***5-8.** In Exercise 5-7, define H_0, H_1, a type I error, and a type II error. Which of these two errors is the more serious error in this case? Explain. Use Table A-2 to test $H_0\ p = \frac{1}{2}$.

***5-9.** In the urban school of Exercise 4-9, eight boys and six girls are to be selected for a color discrimination study. Let X_B equal the number of color-blind boys in the sample and X_G equal the number of color-blind girls. Compute the following probability with the hypergeometric formula and then approximate it with the corresponding binomial formula:

(*a*) $P[(X_B = 2) \cap (X_G = 0)]$
(*b*) Is the approximation reasonable? Why?
(*c*) Use Table A-3 to test for the independence of color blindness with sex, using the following table:

	Boys	*Girls*	*Total*
Color-blind	2	0	2
Not color-blind	6	6	12
Total	8	6	14

***5-10.** Continuous variables can always be dichotomized as follows

$A \quad \{Y \geqslant y_0\}$

$\bar{A} \quad \{Y < y_0\}$

As an example, consider the variable, Y: {heights of college senior men} This variable can be dichotomized by letting $y_0 = 6$ feet, so that A consists of men over 6 feet tall, while $\bar{A}$ consists of men under 6 feet tall. Let X equal the number of men over 6 feet tall in a sample of eight men If, in a certain college, $P(Y \geqslant 6) = .20$, determine

(*a*) $P(X \geqslant 6)$.
(*b*) The probability that no man is over 6 feet tall.
(*c*) The probability that more than half the men are over 6 feet tall.
(*d*) The probability that at least half the men are over 6 feet tall.

6 DISCRETE PROBABILITY DISTRIBUTIONS

We have endeavored to trace the development of ideas about probability, especially in youngsters of school age.

To start with, take a simple experiment testing the notion of statistical distribution We show children aged 10 to 16 a bowl containing blue and yellow beads and inform them that there are equal numbers of blue and yellow beads in the bowl. The experimenter then draws beads from it at random, four at a time, and puts four beads in each of 16 cups. The children are asked to tell how many of the cups will contain respectively. (1) four blue beads, (2) three blue and one yellow, (3) two blue and two yellow, (4) one blue and three yellow, (5) four yellow.

On the basis of such experiments we have found that children apparently progress through four stages. The younger children (around 10) merely guess vaguely that the five possible combinations are not equally likely Those a little more mature realize that the most frequent (or most probable) content of the cups will be two blue and two yellow beads. At the third stage youngsters advance to the conclusions that one blue and three yellow beads will occur as often as one yellow and three blue, and that four blue and four yellow also have equal probabilities. Finally the older children conclude that the combination of one and three is more likely than all four of the same color. These experiments thus show how, with increasing age and experience, uncertain situations are structured in closer and closer accord with the objectivity of mathematical expectation.

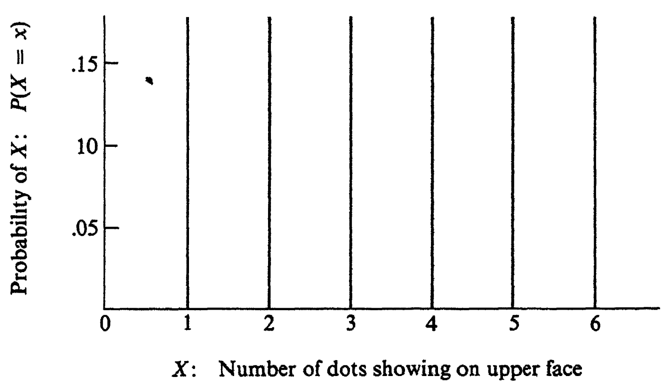
6-1 DISCRETE RANDOM VARIABLES

In Chapter 1, a discrete variable was characterized as a variable that could assume on any one trial a countable set of unique numerical values. For this kind of variable, the possible set of distinct values may be finite or infinite in extent. Examples of discrete variables are the number of correct answers a student selects on a multiple-choice test, the number of correct turns made by a rat in learning a maze on a given run, or the number of farmers who make favorable responses on a questionnaire asking them to express their attitudes toward government farming policies designed to reduce subsidy payments on farm produce. These are examples of discrete variables that take on only a finite number of values. Examples of variables that may take on an infinite number of discrete values are difficult to find in behavioral research. For that reason, most of the discussion presented here and the derivations of important statistical formulas will focus on discrete variables with a finite number of possible values.

Generally, discrete variables are denoted by an X, followed by a listing of the set of distinct values that X can assume on any one trial or single sampling. Thus, for the discrete variable X, one writes $X: \{x_1, x_2, \ldots, x_k\}$. For example, the outcomes on the tossing of a fair die represents a discrete variable X that takes on the set of values $X: \{1,2,3,4,5,6\}$. While this example is neither new nor exceptionally interesting, it should bring to mind other interesting examples of discrete variables discussed in the previous chapters; but, of even greater importance, it should also bring to mind the related and more important concept of a *random variable*, which was also implicitly introduced in earlier pages. This important statistical concept will be studied in considerable detail by means of example, application, and statistical theory. The discussion will begin with the definition of a discrete random variable.

A discrete variable is said to be a discrete random variable if for every value the variable may assume, it is possible to specify the probability that it takes on its set of unique values.

Figure 6-1 Graphic representation of the uniform probability distribution of Table 6-1



Since the probability of any outcome for a fair die is equal to $\frac{1}{6}$, it follows that X, the number of spots showing on the upper face of a tossed die, is a discrete random variable. Furthermore, since the probabilities are all equal, the random variable X is said to have a uniform distribution.

Discrete random variables are usually specified or denoted by means of a set of equations. This is illustrated in Table 6-1 for the tossing of a fair die and its associated discrete variable, X: {number of dots showing on the upper face}. A graphic representation of the probability distribution of this variable is shown in Figure 6-1. Note that in this example the sum of the probabilities is unity. This is true for any discrete random variable.

Table 6-1. Probability distribution for the uniformly distributed random variable X number of dots showing on upper face of a fair six-sided die.

Value of X	*Probability equation*
1	$P(X=1)=\frac{1}{6}$
2	$P(X=2)=\frac{1}{6}$
3	$P(X=3)=\frac{1}{6}$
4	$P(X=4)=\frac{1}{6}$
5	$P(X=5)=\frac{1}{6}$
6	$P(X=6)=\frac{1}{6}$
Total	$\sum_{k=1}^{6} P(X=x_k)=1$

In summary, X is said to be a discrete random variable if

1. For each possible outcome, a probability of occurrence may be associated, such as $P(X=x_k)=p_k$.
2. The sum of the probabilities adds to $1: p_1+p_2+p_3+\cdots+p_K=1$.

Another discrete random variable that was considered in some detail in the previous chapter was the variable X, defined as the number of heads appearing when a coin is tossed N times. To see that this does indeed define a discrete random variable, consider the case where $N=5$. The possible values for X are X: {0,1,2,3,4,5}. As is recalled, the probabilities associated with these events are specified by the binomial formula. These probabilities are shown in Table 6-2. This variable is called a binomial random variable with $N=5$ and $p=\frac{1}{2}$ and is said to have a binomial distribution. The intuitively generated variable of the quotation at the beginning of this chapter is binomial with $N=4$ and $p=\frac{1}{2}$.

Table 6-2. Probability distribution for the binomial random variable X: number of heads appearing in five tosses of a fair coin.

Value of X	*Binomial formula*	*Probability*
0	$\binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5$	$\frac{1}{32}$
1	$\binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4$	$\frac{5}{32}$
2	$\binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3$	$\frac{10}{32}$
3	$\binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2$	$\frac{10}{32}$
4	$\binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1$	$\frac{5}{32}$
5	$\binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0$	$\frac{1}{32}$
Total	$\sum_{x=0}^{5}\binom{5}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{5-x}$	$\frac{32}{32}$

As another example of a discrete random variable, consider a tasting experiment in which a subject is to identify correctly four brands of cigarettes. For any one subject, the number of correctly identified brands is given by X: {0,1,2,3,4}. If the subject just guesses and makes random choices, the probabilities of these outcomes are easily shown to be as reported in Table 6-3. Distributions of this nature are

Table 6-3. Probability distribution for the matching random variable X: number of correctly named brands of four different cigarettes.

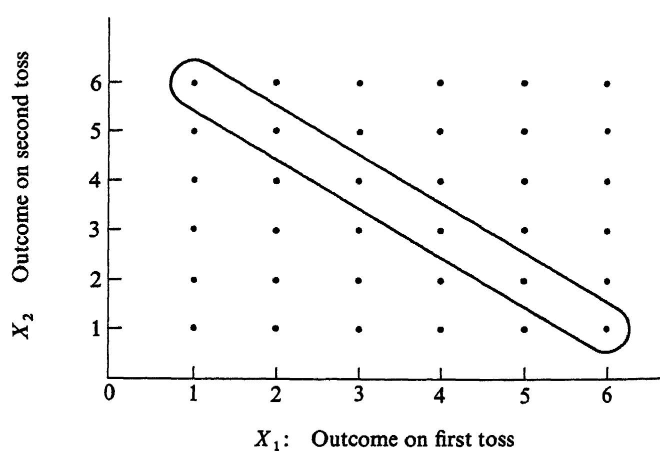
Value of X	$P(X = x_k)$
0	$\frac{9}{24}$
1	$\frac{8}{24}$
2	$\frac{6}{24}$
3	0
4	$\frac{1}{24}$
Total	$\frac{24}{24}$

called matching distributions, since the subject is asked to match the tested products to their brand names. Notice that in this case, as in the binomial example, the probabilities are not equally likely. It should be emphasized that in behavioral research, differential probabilities of outcomes are the usual case. Furthermore, in most experiments and research studies, probabilities are generally unknown. Consequently, the most that a researcher can do is to estimate the probabilities from observed data. In general, they cannot be specified prior to data collection. As a result, one of the major reasons for performing experiments is to estimate unknown probabilities Statistical inference and hypothesis testing are dependent upon good estimates of probability distributions of random variables. For this reason an understanding of probability distributions is of paramount importance in the application of statistical inference methods to hypothesis testing in behavioral research contexts.

6-2 DISCRETE RANDOM VARIABLES GENERATED FROM OTHER DISCRETE RANDOM VARIABLES

From simple random variables one can create complex but interesting random variables whose importance exceeds that of the variables from which they are generated. For example, consider the tossing of a fair die. One could toss the die, observe its outcome, retoss the die, and observe the second outcome. The two outcomes could then be added together. This would produce a new variable that can be denoted by T. If X_1 refers to the outcome on the first toss and X_2 refers to the outcome on the second toss, then $T = X_1 + X_2$. Possible values for T are: T: {2,3,4,5,6,7,8,9,10,11,12}. In this context, the event $T = 2$ is identical to the event $(X_1 = 1 \cap X_2 = 1)$. In like manner, the event $T = 3$ represents either one of

Figure 6-2 The sample spaces of the experiment consisting of two tosses of a fair six-sided die.



the following two events: $(X_1 = 1 \cap X_2 = 2)$ or $(X_1 = 2 \cap X_2 = 1)$. Proceeding in this manner, one can easily determine the probability distribution of T. It is as shown in Table 6-4. The sample space of $X_1 \otimes X_2$ is shown in Figure 6-2. Since the probabilities of T are not uniform, the outcomes are not equally likely. The event $T = 7$

Table 6-4. Probability distribution of the triangular random variable $T = X_1 + X_2$: total number of dots appearing on two tosses of a fair six-sided die.

Value of T	$P(T = t_k)$
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$
Total	$\frac{36}{36}$

has the greatest probability, while the events $T = 2$ or $T = 12$ have the smallest probability of occurrence. In Figure 6-2 the event $T = 7$ consists of the six circled points along the main diagonal. The remaining outcomes that give rise to T are represented along the remaining ten diagonal lines parallel to the main diagonal. The distribution of T is called the triangular distribution.

As another example of a random variable that can be generated on the sample space illustrated in Figure 6-2, consider the variable $U = (X_1 - 3.5)^2 + (X_2 - 3.5)^2$. This rather complex-appearing variable will play a major role in the statistical theory of variability developed in the succeeding pages. For the outcome (1,1), $U = (1 - 3.5)^2 + (1 - 3.5)^2 = (-2.5)^2 + (-2.5)^2 = 6.25 + 6.25 = 12.50$; for the outcome (1,2), $U = (1 - 3.5)^2 + (2 - 3.5)^2 = (-2.5)^2 + (-1.5)^2 = 6.25 + 2.25 = 8.50$; . . ; and, finally, for the outcome (6,6), $U = (6 - 3.5)^2 + (6 - 3.5)^2 = (2.5)^2 + (2.5)^2 = 6.25 + 6.25 = 12.50$. The entire set of possible values for U along with their associated probabilities is shown in Table 6-5. As can be seen, the probability distribution of U is quite irregular. For discrete variables, this irregularity is not unusual. Note that this new distribution was generated quite simply from the uniform distribution. What is perhaps even more surprising is that discrete random variables with uniform

Table 6-5. Probability distribution of the random variable $U = (X_1 - 3.5)^2 + (X_2 - 3.5)^2$, where X_1 and X_2 represent the number of dots appearing on the upper faces of two six-sided dice.

Value of U	$P(U = u_k)$
.50	$\frac{4}{36}$
2.50	$\frac{8}{36}$
4.50	$\frac{4}{36}$
6.50	$\frac{8}{36}$
8.50	$\frac{8}{36}$
12.50	$\frac{4}{36}$
Total	$\frac{36}{36}$

distribution can generate random variables that have binomial distributions. As an example of how this can happen, consider the random variable X: the number of times a 6 appears on a die that is tossed four times. For this variable, $p = \frac{1}{6}$ and $q = \frac{5}{6}$. Possible values are X: {0,1,2,3,4}. Thus, the probability distribution of X can be summarized as shown in Table 6-6. Notice that the appearance of three or four 6's in four tosses is very rare. If a die were tossed four times and three or four

Table 6-6. Probability distribution for the binomial random variable X: number of 6's that appear in four tosses of a fair six-sided die.

Value of X	*Binomial formula*	*Probability*
0	$\binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4$	.4823
1	$\binom{4}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^3$	.3858
2	$\binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2$	.1157
3	$\binom{4}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^1$	.0154
4	$\binom{4}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^0$	.0008
Total	$\sum_{x=0}^{4}\binom{4}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{4-x}$	1.0000

6's were observed, there might be good reason to doubt the fairness either of the die or of the tosses.

As might be suspected, tables representing probability distributions become extremely unwieldy to handle as the number of possible values of X increases and so summary procedures become important and desirable. There are two major summary procedures used by researchers: graphic methods and numerical methods. The easiest to understand and use are the graphic methods. Some graphic procedures appropriate for discrete random variables are illustrated in the following section.

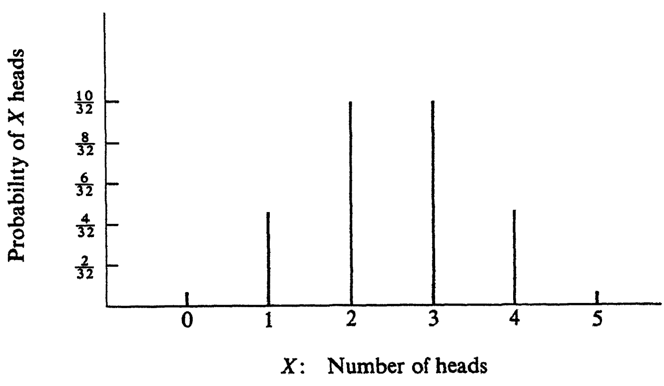
6-3 GRAPHIC SUMMARY PROCEDURES FOR DISCRETE RANDOM VARIABLES

For discrete random variables there are a number of graphic methods in use today. The two most frequently used are the line graph and the cumulative probability polygon.

A line graph shows the distribution characteristics or features of discrete random variables. While cumulative probability polygons are harder to interpret, they are perhaps the graph with the greater utility since descriptive summary measures can be obtained from them with considerable ease.

Both graphic methods will be illustrated for the binomial probability distribution shown in Table 6-2, which represents the number of heads that appear when one coin is tossed five times or when five identical coins are tossed simultaneously. In a line graph, the probabilities of each value of X are represented by a line equal in length to the probabilities of X and drawn parallel to the ordinate or vertical axis above the values of X, which are listed in their natural numerical order on the horizontal axis or abscissa of the graph. Both scales are then carefully labeled and the graph is given a name so that it can be removed from context and anyone may read it and interpret it. If a graph cannot be read without recourse to the text, it is a useless graph. The line graph for the probability distribution of Table 6-2 is shown in Figure 6-3.

Figure 6-3. Graphic representation of the binomial random variable X number of heads appearing when a fair coin is tossed five times.



For emphasis, note that for each value of X a line is drawn perpendicular to the horizontal axis with a height exactly equal to the probability of occurrence. This means that the sum of the lengths of the line segments is equal to unity. In this particular example, the graph shows that the probability distribution of X is balanced and is symmetric about $X = 2.5$. Probability distributions that are symmetric about their midpoint are said to be symmetric distributions.

The cumulative probability polygon is a little more difficult to construct and a bit harder to interpret. For this graph, one considers the cumulative set of events: $\{X \leqslant x_1, X \leqslant x_2, X \leqslant x_3, \ldots, X \leqslant x_K\}$. For this example, the probability that $X \leqslant 0$ is $\frac{1}{32}$. The probability that $X \leqslant 1$ is $\frac{6}{32}$, since $P(X \leqslant 1) = P(X = 0) + P(X = 1) = \frac{1}{32} + \frac{5}{32} = \frac{6}{32}$. In like manner, the probability that $X \leqslant 2$ is $\frac{16}{32}$; the probability that $X \leqslant 3$ is $\frac{26}{32}$; the probability that $X \leqslant 4$ is $\frac{31}{32}$; and the probability that $X \leqslant 5$ is 1. These are called cumulative probabilities. These probabilities can be summarized in table form, as shown in Table 6-7. The cumulative probability polygon constructed from this table is shown in Figure 6-4.

For this graph note that

$$\begin{aligned} P(X < 2) &= P(X \leqslant 1.76) = P(X \leqslant 1.369241) \\ &= P(X \leqslant 1.00002) = .18750 \end{aligned}$$

but as soon as $X < 1$, the probability drops to .03125. Graphs with this property are frequently referred to as step functions. They are appropriate for discrete variables.

Figure 6-4 Cumulative probability distribution of the binomial random variable X number of heads appearing when a fair coin is tossed five times

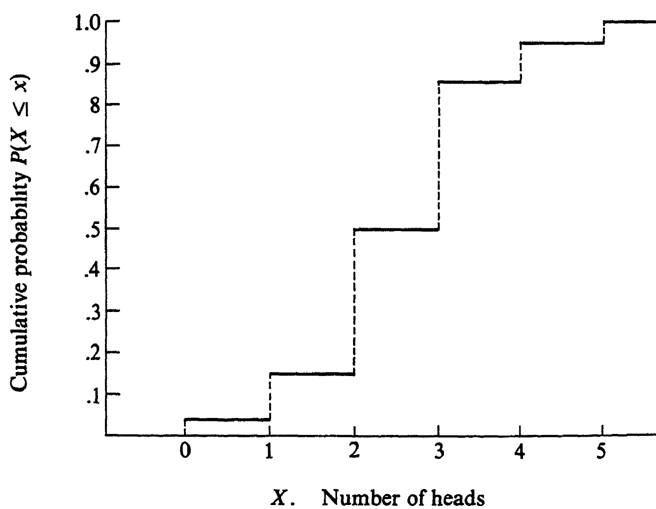


Table 6-7. Cumulative probability distribution of the binomial random variable X. number of heads that appear when a fair coin is tossed five times.

Value of X	*Cumulative probability* $P(X \leqslant x_k)$
0	$\frac{1}{32}$ = 03125
1	$\frac{6}{32}$ = .18750
2	$\frac{16}{32}$ = .50000
3	$\frac{26}{32}$ = .81250
4	$\frac{31}{32}$ = 96875
5	$\frac{32}{32}$ = 1.00000

Because graphs can be constructed to deceive a reader by a simple stretch or compression of the length of the scales, they are not very useful for analytical purposes—they are useful only for descriptive purposes. However, the cumulative probability polygon does not possess this disadvantage to the same degree because the vertical scale extends from 0 to 1. Even though this scale can be elongated or condensed to influence visual impressions, the vertical axis of this graph can always be read in an invariant manner no matter how much distortion has been forced into the graph because of distortion of the scale.

One other way of summarizing the information contained in a probability distribution is to specify numbers that convey pertinent information. Perhaps the most important of these are numbers that convey information about the center and the variability about the center. Four useful measures related to the middle and the extent of a probability distribution are its median and expected value and its variance and standard deviation. These measures contain summary information about the center and the spread of the probability along the set of possible outcomes.

6-4 THE MEDIAN VALUE OF A DISCRETE RANDOM VARIABLE

For the binomial variable used to illustrate the two graphic methods of presentation for discrete random variables, it was noted that X is symmetrical about its midpoint value of 2.5. This means

$$P(X \leqslant 2.5) = P(X \geqslant 2.5) = \tfrac{1}{2}$$

The number M, which has the property that $P(X \leqslant M) = P(X \geqslant M) = \frac{1}{2}$, is said to be the median of the distribution of X. Thus, $M = 2.5$ is the *median* value of X. However, from the step function it is also seen that any value of X in the interval $2 \leqslant X < 3$ is also a median. As this suggests, discrete random variables may have

an infinite set of medians. Since behavioral scientists prefer to have only one value M satisfying the relation $P(X \leqslant M) = P(X \geqslant M) = \frac{1}{2}$, it is convenient to define the midpoint of the interval that contains the set of medians of a discrete random variable as the median.

As an example, consider the probability distribution of the outcomes on the tossing of a fair die. Examination of Figure 6-1 suggests that the median is equal to 3.5. However, according to the definition of the median, the range from $3.0 \leqslant X < 4$ must be considered as representing the median of the distribution since all of these numbers satisfy the definition of the median. Thus, the median for this distribution is not unique. In this case one can say that there is no median or that there is an infinite number of them. However, by the midpoint convention, a unique median can be obtained by defining the median value equal to the midpoint of the interval in which the infinite set of medians exists. In this case, $M = 3.5$.

6-5 THE EXPECTED VALUE OR POPULATION MEAN OF A DISCRETE RANDOM VARIABLE

The expected value or population mean of a discrete random variable is a much more useful measure of center than is the median, but it is a bit harder to comprehend. Consider a gambler who plays a game that costs \$1 per trial. The game consists of a random selection of a card from a deck of 52 cards. If the gambler selects the ace of spades he wins \$20 plus his \$1 fee; if he selects a face card he wins \$2 plus his \$1 fee; and if he selects any other card, he loses his dollar. If he plays the game one time, what is his expected gain? To answer this question, note that $P(\text{ace of spades}) = \frac{1}{52}$, $P(\text{face card}) = \frac{12}{52}$, and $P(\text{losing card}) = \frac{39}{52}$. In every 52 trials he has 1 chance to gain \$20, 12 chances to gain \$2, and 39 chances to lose \$1. Thus, in 52 trials he should earn about

$$(1)(\$20) + (12)(\$2) + (39)(-\$1) = \$5$$

so that his average gain per trial should be about

$$\$\tfrac{5}{52} = \$.096$$

Combine these two results to find his expected gain per trial, given by

$$\text{Expected gain} = \$20(\tfrac{1}{52}) + \$2(\tfrac{12}{52}) + (-\$1)(\tfrac{39}{52}) = \$\,096$$

Thus, to obtain the expected gain, one should multiply the possible payoffs by their probabilities and add the resulting products. In this case, this gives an expected payoff of \$.096 per game. Without doubt he should be encouraged to play this game as often as he can, for in each 1,000 trials he can expect to win \$96. This simple example is based on the following definition of the *expected value* of a discrete random variable.

The expected value of a discrete random variable X is equal to the sum of the products of all possible values that the variable can assume, times the probability that it takes on those values. Symbolically, this is written as

$$E(X) = \sum_{k=1}^{K} x_k p_k = \sum_{k=1}^{K} x_k P(X = x_k)$$

$$= x_1 P(X = x_1) + x_2 P(X = x_2) + x_3 P(X = x_3) + \cdots + x_K P(X = x_K)$$

The use of this definition is illustrated for X: the number of brands of cigarettes correctly identified by a tested subject. By definition,

$$E(X) = (0)(\tfrac{9}{24}) + (1)(\tfrac{8}{24}) + (2)(\tfrac{6}{24}) + (3)(0) + (4)(\tfrac{1}{24})$$
$$= \tfrac{24}{24} = 1$$

Since a subject must identify 0, 1, 2, 3, or 4 brands, one may wonder how to interpret $E(X) = 1$. The interpretation of this number depends upon the relative-frequency interpretation of probability. Think of the experiment as being repeated many times on similar subjects with the same skill in identifying cigarette brands. The number $E(X) = 1$ represents the average number of cigarettes one subject will identify by guessing. Thus, if 1,000 subjects are each given four cigarettes to identify, one expects them to make 1,000 correct identifications out of 4,000 cigarettes tested. Some testers may make four correct identifications, and some will make none. If their total experiences are combined and if the average number of correct identifications is computed, this average will be close in numerical value to $E(X) = 1$, provided that they guess. If they really know the brands, then one would expect $E(X) > 1$. In any case, the expected value of a discrete random variable is synonymous with its long-term-average value.

When the possible set of X values is large, it is usually easier to compute $E(X)$ in table form, as illustrated in Table 6-8. In this case, $E(X) = \frac{24}{24} = 1$.

Table 6-8. Computations for determining the expected number of cigarette brands correctly identified by an individual who tests four brands by just guessing.

Value of X	*Probability of X* $P(X = x_k)$	$x_k P(X = x_k)$
0	$\frac{9}{24}$	0
1	$\frac{8}{24}$	$\frac{8}{24}$
2	$\frac{6}{24}$	$\frac{12}{24}$
3	0	0
4	$\frac{1}{24}$	$\frac{4}{24}$
Total	$\frac{24}{24}$	$\frac{24}{24}$

Since the expected value of a probability distribution is one of the basic notions of statistics, a special notation has evolved to represent $E(X)$. It is the Greek letter μ (mu). It will be used interchangeably in this text with $E(X)$.

As another example, consider the expected value for the tossing of a fair die. Computations leading to its expected value are shown in Table 6-9. As the arithmetic

Table 6-9. Computations for determining the expected number of dots showing on the upper face of a fair six-sided die.

Value of X	$P(X=x_k)$	$x_k P(X=x_k)$
1	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{2}{6}$
3	$\frac{1}{6}$	$\frac{3}{6}$
4	$\frac{1}{6}$	$\frac{4}{6}$
5	$\frac{1}{6}$	$\frac{5}{6}$
6	$\frac{1}{6}$	$\frac{6}{6}$
Total	1	$\frac{21}{6}$

shows, when a fair die is tossed the expected number of dots is given by $E(X)=\frac{21}{6}=3.5$. In this case, the expected value is not included in the set of possible values. This appears to be a contradiction of meaning. However, it is not if the relative-frequency interpretation of probability is extended to this case. As with the previous example, $E(X)$ can be interpreted as the long-term-average outcome over a long series of trials. On any one trial, possible outcomes are X: $\{1,2,3,4,5,6\}$. These are expected with probability $\frac{1}{6}$. If the die is tossed 600 times, 100 1's, 100 2's, . . . , are expected. The average value of all 600 values should be close to 3.5, the average value of an infinite number of trials. The example also shows that if a probability distribution is symmetrical, its expected value will be equal to the midpoint of the range of the variable, which in turn will equal the median.

As another example, consider the sum that appears when two dice are tossed. Let X_1 be the number showing on the first die and let X_2 be the number showing on the second die. The sum can be represented by the variable $T=X_1+X_2$. The probability distribution of T and the computations leading to the determination of the expected value are shown in Table 6-10. The expected value is given by $E(T)=\frac{252}{36}=7$. Of course, this is reasonable, since as already shown, the $E(X_1)=3.5$ and the $E(X_2)=3.5$. Intuition should suggest that $E(T)=7=3.5+3.5=E(X_1)+E(X_2)$. In a similar manner, if a die is tossed three times, the expected value should be 10.5. The expectation for the first toss is 3.5, and it is also 3.5 on the second and 3.5 on the third. Thus, the expected value for the sum is 10.5. Note that the expected value does not occur in any of these cases. At no time has it been stated that a variable

Table 6-10. Computations for determining the expected number of dots showing on the upper faces of two simultaneously tossed fair six-sided dice.

Value of T	$P(T=t_k)$	$t_k P(T=t_k)$
2	$\frac{1}{36}$	$\frac{2}{36}$
3	$\frac{2}{36}$	$\frac{6}{36}$
4	$\frac{3}{36}$	$\frac{12}{36}$
5	$\frac{4}{36}$	$\frac{20}{36}$
6	$\frac{5}{36}$	$\frac{30}{36}$
7	$\frac{6}{36}$	$\frac{42}{36}$
8	$\frac{5}{36}$	$\frac{40}{36}$
9	$\frac{4}{36}$	$\frac{36}{36}$
10	$\frac{3}{36}$	$\frac{30}{36}$
11	$\frac{2}{36}$	$\frac{22}{36}$
12	$\frac{1}{36}$	$\frac{12}{36}$
Total	$\frac{36}{36}$	$\frac{252}{36}$

will take on its expected value. The expected value is an algebraically defined number that essentially describes the numerical center of the probability distribution of a variable.

As another example, consider tossing a fair coin five times; the probability distribution of heads is binomial with $N=5$ and $p=\frac{1}{2}$. By definition,

$$E(X)=\sum_{k=0}^{5} x_k P(X=x_k)=(0)(\tfrac{1}{32})+(1)(\tfrac{5}{32})+(2)(\tfrac{10}{32})+(3)(\tfrac{10}{32})$$
$$+(4)(\tfrac{5}{32})+(5)(\tfrac{1}{32})$$
$$=\tfrac{80}{32}=2.5$$

In five tosses, heads are expected to appear 2.5 times, provided that the coin is fair and the tosses are independent. Again, there is no statement in the definition that the expected value is a value that can occur at any one trial.

Suppose the coin is tossed 10 times. One might wonder what the expected number of heads in 10 tosses will equal. To determine this one could consider the experiment as taking place in two stages. The first stage consists of the first 5 tosses while the second stage would consist of the second 5 tosses. Thus, one would expect that for the total number of heads, $T=X_1+X_2$. The total number of heads expected should be given by

$$E(T)=E(X_1)+E(X_2)=2.5+2.5=5$$

If the arithmetic were to be completed via the defining equation for expected values, it would be seen that $E(T)$ is indeed equal to 5.

These last examples illustrate an important property of expected values, namely that the expected value of the sum of two random variables is equal to the sum of their expected values. This is a special case of the following important theorem, which is stated without proof.

Theorem 6-1. Expected value of a sum of random variables

Suppose that $T = X_1 + X_2 + X_3 + \cdots + X_N$; then the expected value of T is given by

$$E(T) = E(X_1) + E(X_2) + E(X_3) + \cdots + E(X_N)$$

If all the X's have the same probability distribution, then

$$E(X_1) = E(X_2) = E(X_3) = \cdots = E(X_N) = \mu$$

so that

$$E(T) = N\mu$$

6-6 THE EXPECTED VALUE OF A BINOMIAL RANDOM VARIABLE

As a special case of Theorem 6-1, it follows that the expected value of a binomial random variable is given by $E(T) = Np$. This is stated as Theorem 6-2.

Theorem 6-2

If T is a binomial random variable with N trials and probability of success p, then

$$E(T) = Np$$

Proof. Consider the possible set of outcomes on the first trial If a success occurs, record a 1 and if a failure occurs, record a 0. As a result, $P(X = 1) = p$ and $P(X = 0) = q$ Thus, by definition,

$$\begin{aligned} E(X) &= \sum_{k=0}^{1} x_k P(X = x_k) \\ &= (0)P(X = 0) + (1)P(X = 1) \\ &= (0)q + (1)p \\ &= p \end{aligned}$$

For N trials,

$$T = X_1 + X_2 \cdots + X_N$$

but since all trials are identical, it follows by Theorem 6-1 that

$$E(X_1) = E(X_2) = \cdots = E(X_N) = p$$

so that

$$\begin{aligned} E(T) &= N\mu \\ &= Np \end{aligned}$$

This completes the proof

As an example, consider a single item on a true-false questionnaire. Responses to this single item are either correct or wrong. Every time a correct answer is made a score of +1 is assigned. Every time a wrong answer is given a score of 0 is assigned. If the probability of a correct response is equal to p, the probability of a wrong response is $q = 1 - p$. Thus, X: {0,1} is a binomial variable. The expected value of this variable is

$$E(X) = \sum_{k=0}^{1} x_k P(X = x_k) = (0)q + (1)p = p$$

If the exam contains 100 items and if the probability of a correct response to each item is p, then the expected score on the test, according to Theorem 6-2, is

$$E(T) = (100)p$$

If the probability of getting each of the individual items correct is .8, then the expected score is 100 times .8, or 80. If the test has N items, $E(T) = Np$. This is the expected value for any binomial variable. As shown earlier, the expected number of heads appearing when a fair coin is tossed five times is equal to 2.5. With this last formula, it is seen that the expected value is given by the product $Np = 5(\frac{1}{2}) = 2.5$. As another example, consider a city with one million people for which the probability of being hit and killed by an automobile in a particular year is .001. In one year, $Np = 1{,}000{,}000(.001) = 1{,}000$ people are expected to be killed by an automobile.

6-7 LONG-RUN INTERPRETATION OF THE EXPECTED VALUE OF A RANDOM VARIABLE

The expected value of average population value of a discrete random variable has been defined as

$$E(X) = \sum_{k=1}^{K} x_k P(X = x_k)$$

One way to interpret the expected value of a random variable is in terms of the long-run relative-frequency interpretation of probability. Imagine an experiment that is repeated over and over under identical conditions. As is known, the outcomes of these experiments will still vary; it is impossible to keep the environment always the same. If the experiment were repeated an exceptionally large number of times and a simple average computed over them, the derived number would be the expected value. In this sense, the expected value of a random variable is the long-term average of a repeated experiment. Of course, it should be realized that in practice an experiment is done once or perhaps twice. However, in theory, the experiment could be repeated many times.

6-8 SKEWED-PROBABILITY DISTRIBUTION

If a probability distribution is not symmetrical, it is said to be *skewed.* If the *tail* of the probability distribution extends along the positive direction of the X axis, the distribution is said to be *positively skewed.* If the *tail* extends along the negative direction, the distribution is said to be *negatively skewed.* Examples of positively and negatively skewed distributions are shown in Figures 6-5 and 6-6.

If a distribution is symmetrical, the expected value and the median are equal; therefore, if a distribution is skewed, then one of these measures of central tendency must be larger than the other. In general, the expected value is pulled toward the tail. Thus, for a positively skewed distribution, $M < E(X)$, and for a negatively skewed distribution, $M > E(X)$.

6-9 THE MODE OF A PROBABILITY DISTRIBUTION

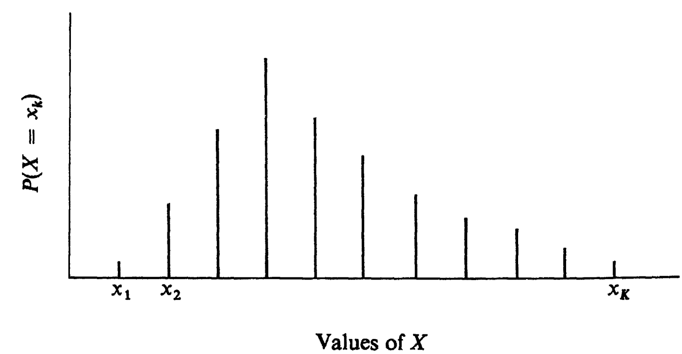
Another measure of center is the mode of the probability distribution. The *mode* ($\mathscr{M}$) is that value of X for which the probability is greatest. For the most part, the mode is a rather useless measure of center since it is almost never used to argue from some to all.

6-10 VARIANCE OF A PROBABILITY DISTRIBUTION

From the examples discussed in previous sections, it is quite apparent that some probability distributions have a narrow range of values while others are quite extreme in their spread. While many measures have been proposed to reflect differences in variability, the most frequently employed measure of variation used in statistics is called the variance of a probability distribution. This indicator of variability measures deviation from expectation.

To help define and motivate the use of the variance as a measure of variability, consider each value that a discrete variable can assume: $\{x_1, x_2, \ldots, x_K\}$. With μ as a center of reference, $\{x_1 - \mu, x_2 - \mu, \ldots, x_K - \mu\}$ measures how much each value of X differs from the expected value of the variable. These individual deviations

Figure 6-5 Illustration of a positively skewed discrete random variable.



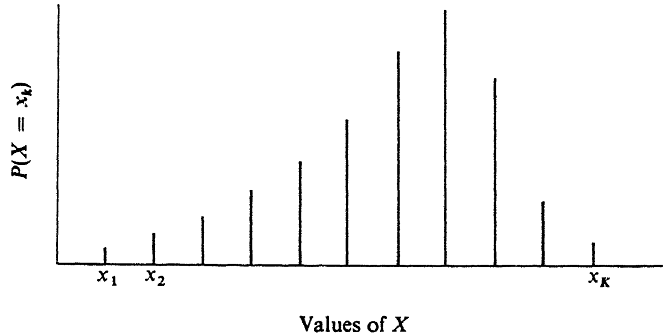


Figure 6-6. Illustration of a negatively skewed discrete random variable

measure the total variation in X. To create a summary measure, it might seem that weighting each deviation by its probability and then adding across all deviations would produce a measure of variance. While this is intuitively plausible, it is not very useful since this sum will reduce to 0.

This last property can be stated as a theorem.

Theorem 6-3

The sum of the weighted deviations about the expected value of a discrete random variable is equal to 0.

Proof. Consider each deviation multiplied by its probability and summed. Thus

$$\begin{aligned} Q &= (x_1 - \mu)p_1 + (x_2 - \mu)p_2 + \cdots + (x_K - \mu)p_K \\ &= (x_1p_1 + x_2p_2 + \cdots + x_Kp_K) - \mu(p_1 + p_2 + \cdots + p_K) \end{aligned}$$

By definition,

$$\mu = x_1p_1 + x_2p_2 + \cdots + x_Kp_K$$

so that

$$Q = \mu - \mu(p_1 + p_2 + \cdots + p_K)$$

Since X has a probability distribution

$$p_1 + p_2 + \cdots + p_K = 1$$

then

$$Q = \mu - \mu(1) = \mu - \mu = 0$$

This completes the proof.

Yet, it still makes sense to define the variable in terms of these deviations. The difficulty in doing this effectively arises because the positive deviations balance the negative deviations. To get around this one could square the deviations, sum across

them, and then take the square root so as to get back to the original scale. If this is acceptable then the *variance* should be defined as follows

$$\begin{aligned}\mathrm{Var}(X) &= (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + \cdots + (x_K - \mu)^2 p_K \\ &= \sum_{k=1}^{K} (x_k - \mu)^2 p_k\end{aligned}$$

As indicated, one need only take the square root of the resulting number to get back to the original scale. This square root is called the standard deviation of the random variable X and is generally denoted by the Greek letter σ (sigma.) Thus, the *standard deviation* of X is defined as follows:

$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\sum_{k=1}^{K} (x_k - \mu)^2 p_k}$$

Sometimes the variance will be denoted by

$$\mathrm{Var}(X) = \sum_{k=1}^{K} [x_k - E(X)]^2 P(X = x_k)$$

For the most part, the variance is a measure that is extremely difficult to interpret. However, when comparing the variability of two probability distributions it is much easier to interpret.

To aid in understanding what the variance measures, consider the two distributions shown in Figures 6-1 and 6-3. Note that in the range of values for the uniform distribution of Figure 6-1 the range is given by $R_U = X_6 - X_1 = 6 - 1 = 5$, while for the binomial distribution of Figure 6-3, the range is given by $R_B = X_5 - X_0 = 5 - 0 = 5$. Both distributions have equal ranges. Even though the range of outcomes is the same, the distribution of Figure 6-3 shows much less variability than the distribution of Figure 6-1. For example, the probabilities of the two central values for these distributions are as follows:

1. For the binomial distribution,

   $$P(X = 2) + P(X = 3) = \tfrac{10}{32} + \tfrac{10}{32} = \tfrac{20}{32} = .625$$

2. For the uniform distribution,

   $$P(X = 3) + P(X = 4) = \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{2}{6} = .333$$

On this basis alone, one would expect the variance of the distribution of Figure 6-3 to be less than the variance of the distribution of Figure 6-1. That this is true will now be shown.

For the uniform distribution with possible outcomes X: $\{1,2,3,4,5,6\}$ with probabilities all equal to $\frac{1}{6}$, the variance is computed as shown in Table 6-11. The variance of X is given by $\sigma^2 = 17\ 5/6 = 2\ 9167$ and the standard deviation is given by $\sigma = \sqrt{2.9167} = 1.708$.

Table 6-11. Computations for determining the variance of the uniform random variable X: number of dots showing on upper face of a fair six-sided die.

Value of X	$P(X = x_k)$	$x_k - E(X)$	$[x_k - E(X)]^2$	$[x_k - E(X)]^2 P(X = x_k)$
1	$\frac{1}{6}$	−2.5	6.25	6.25/6
2	$\frac{1}{6}$	−1.5	2.25	2.25/6
3	$\frac{1}{6}$	−.5	.25	.25/6
4	$\frac{1}{6}$	.5	.25	.25/6
5	$\frac{1}{6}$	1.5	2.25	2.25/6
6	$\frac{1}{6}$	2.5	6.25	6.25/6
Total				17.5/6

Corresponding computations for the binomial distribution are summarized in Table 6-12. For this variable, the variance of X is given by $\sigma^2 = \frac{40}{32} = 1.25$ and the standard deviation is given by $\sigma = \sqrt{1.25} = 1.11$. Accordingly, the variance of the uniform distribution with the same range is larger. The ratio of the two variances is 2.9167 to 1.25 or 2.2 to 1. Thus, even though the ranges for the two distributions are identical, the variability for the uniform distribution is much more than twice that of the binomial distribution.

Table 6-12. Computations for determining the variance of the binomial random variable X: number of heads appearing in five tosses of a fair coin.

Value of X	$P(X = x_k)$	$x_k - E(X)$	$[x_k - E(X)]^2$	$[x_k - E(X)]^2 P(X = x_k)$
0	$\frac{1}{32}$	−2.5	6.25	6.25/32
1	$\frac{5}{32}$	−1.5	2.25	11.25/32
2	$\frac{10}{32}$	−.5	.25	2.50/32
3	$\frac{10}{32}$	.5	.25	2.50/32
4	$\frac{5}{32}$	1.5	2.25	11.25/32
5	$\frac{1}{32}$	2.5	6.25	6.25/32
Total				40/32

For the most part, it can be said that knowledge concerning the value of a population variance is not of great importance, since the interpretation of this measure is not simple. However, its usefulness in comparing the variability between two or more distributions is of considerable value. For example, it is known that the binomial variable of the previous example fluctuates less about its central value than does the uniform variable about its center. In addition, as will be seen in

Chapter 7, the square root of the variance, or the standard deviation, has considerable value in describing variability for certain continuous variables. Finally, the importance of variance as a descriptive measure of variability is embodied in the theorem following the next example. As will be seen, variances can be added. The same is not true of standard deviations; standard deviations can never be added to produce another standard deviation. For these reasons, variance has greater theoretical importance than does the standard deviation, even though the standard deviation is easier to interpret.

As another example, consider the probability distribution of Table 6-10, which is the distribution of $T = X_1 + X_2$ where X_1 and X_2 are the outcomes on the tossing of two independent dice. As is recalled, it was seen that $E(T) = E(X_1) + E(X_2)$. In like manner, one might suspect that $\mathrm{Var}(T) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$. That this is, indeed, true is shown in Table 6-13.

Table 6-13. Computations for determining the variance of the triangular random variable $T = X_1 + X_2$: total number of dots appearing on two tosses of a fair six-sided die.

Value of T	$P(T = t_k)$	$t_k - E(T)$	$[t_k - E(T)]^2$	$[t_k - E(T)]^2 P(T = t_k)$
2	$\frac{1}{36}$	−5	25	$\frac{25}{36}$
3	$\frac{2}{36}$	−4	16	$\frac{32}{36}$
4	$\frac{3}{36}$	−3	9	$\frac{27}{36}$
5	$\frac{4}{36}$	−2	4	$\frac{16}{36}$
6	$\frac{5}{36}$	−1	1	$\frac{5}{36}$
7	$\frac{6}{36}$	0	0	0
8	$\frac{5}{36}$	1	1	$\frac{5}{36}$
9	$\frac{4}{36}$	2	4	$\frac{16}{36}$
10	$\frac{3}{36}$	3	9	$\frac{27}{36}$
11	$\frac{2}{36}$	4	16	$\frac{32}{36}$
12	$\frac{1}{36}$	5	25	$\frac{25}{36}$
Total				$\frac{210}{36}$

The variance on the tossing of two dice simultaneously, or two independent tosses of the same die, is given by

$$\sigma_T^2 = \tfrac{210}{36} = 5.8333$$

From this it is also easy to see that

$$\begin{aligned}\mathrm{Var}(T) &= 5.8334 = 2.9167 + 2.9167\\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2)\end{aligned}$$

This indicates that variances add in the same manner as do expected values. Like the statement on expected values, this is also a special case of the following theorem, which is presented without proof.

Theorem 6-4

Suppose that $T = X_1 + X_2 + \cdots + X_N$ and assume that the variables are *statistically independent*. Then the variance of T is given by

$$\mathrm{Var}(T) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_N)$$

If all the X's have the same probability distribution, then

$$\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \cdots = \mathrm{Var}(X_N) = \sigma_X^2$$

so that

$$\mathrm{Var}(T) = N\sigma_X^2$$

6-11 VARIANCE OF A BINOMIAL VARIABLE

As shown earlier, the expected value of a binomial variable with $P(X=0)=q$ and $P(X=1)=p$ is given by $E(X)=p$, and if there are N trials $E(T)=Np$. As might be suspected, the variance of a binomial random variable can be stated in terms of N and p. This relationship is specified in Theorem 6-5.

Theorem 6-5

The variance of a binomial random variable is given by

$$\mathrm{Var}(T) = Np(1-p) = Npq$$

Proof. Consider the set of outcomes on the first trial: $P(X=1)=p$ and $P(X=0)=q$. By definition,

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{k=0}^{1} [x_k - E(X)]^2 P(X = x_k) \\
&= (0-p)^2 q + (1-p)^2 p \\
&= p^2 q + q^2 p \\
&= pq(p+q) \\
&= pq
\end{aligned}$$

Since the trials are identical, $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \cdots = \mathrm{Var}(X_N) = pq$ and

$$\begin{aligned}
\mathrm{Var}(T) &= N\sigma^2 \\
&= Npq \\
&= Np(1-p)
\end{aligned}$$

This completes the proof.

In the example of Table 6-12, $N = 5$, $p = \frac{1}{2}$, and $q = \frac{1}{2}$, so that

$$\mathrm{Var}(T) = (5)(\tfrac{1}{2})(\tfrac{1}{2}) = 1.25$$

exactly as found by means of the defining formula.

6-12 THE EXPECTED VALUE AND VARIANCE OF A HYPERGEOMETRIC VARIABLE

As was seen in Section 5-7, hypergeometric probabilities could be simply approximated by binomial probabilities, provided that the ratio of the sample size to the universe size is less than 10 percent. As a result, one might suspect that the expected value and variance of a hypergeometric distribution should be the same as those of its related binomial distribution. For the expected values, this is true, but for the variance it is false. The appropriate summary measures for hypergeometric variables are stated in the following theorem, which is given without proof.

Theorem 6-6

If a finite universe of size N consists of $N(A)$ elements with property A and $N(\bar{A})$ elements with property $\bar{A}$ and if n observations are taken from this universe, then the variable X: {number of A's in the sample} has a hypergeometric distribution with

$$E(X) = n\frac{N(A)}{N} = np$$

$$\mathrm{Var}(X) = n\frac{N(A)}{N}\frac{N(\bar{A})}{N}\left(\frac{N-n}{N-1}\right)$$

$$= npq\left(\frac{N-n}{N-1}\right)$$

where $p = N(A)/N$ and $q = N(\bar{A})/N$.

As an example of the use of this theorem, consider a survey in which a questionnaire is to be distributed to a random sample of 80 union members from a population of 200 members. If the union consists of 30 foreign-born nationals, then the expected number of foreign-born nationals in the sample is given by

$$E(X) = n\frac{N(A)}{N} = 80\,(\tfrac{30}{200}) = 12$$

and the variance is given by

$$\mathrm{Var}(X) = n\,\frac{N(A)}{N}\,\frac{N(\bar{A})}{N}\left(\frac{N-n}{N-1}\right)$$

$$= 80\left(\frac{30}{200}\right)\left(\frac{170}{200}\right)\left(\frac{200-80}{200-1}\right)$$

$$= 80\,(\tfrac{3}{20})\,(\tfrac{17}{20})\,(\tfrac{120}{199})$$

$$= 6.18$$

The standard deviation is equal to $\sqrt{6.18} = 2.47$. Note that if the binomial theory were incorrectly applied to this model, one would find that

$$\text{Var}(X) = npq = 80\left(\tfrac{30}{200}\right)\left(\tfrac{170}{200}\right) = 10.2$$

which is much too large. If $n/N < 10$ percent, the discrepancy would be considerably less.

6-13 THE COMPUTING FORMULA FOR THE VARIANCE OF A DISCRETE RANDOM VARIABLE

In general, it is time-consuming to determine the variance of a population by means of the defining formula. With a little algebra, it is possible to derive a computing formula that makes the determination of the variance simpler. This formula is stated in Theorem 6-7.

Theorem 6-7

The computing formula for the variance of a discrete random variable is given by

$$\begin{aligned}\text{Var}(X) &= \sum_{k=1}^{K} x_k^2 p_k - \left(\sum_{k=1}^{K} x_k p_k\right)^2 \\ &= E(X^2) - E(X)^2\end{aligned}$$

Proof. The formula for the variance is given by

$$\begin{aligned}\text{Var}(X) &= \sum_{k=1}^{K} (x_k - \mu)^2 P(X = x_k) \\ &= \sum_{k=1}^{K} (x_k - \mu)^2 p_k\end{aligned}$$

According to binomial theory,

$$(x_k - \mu)^2 = x_k^2 - 2\mu x_k + \mu^2$$

so that

$$\begin{aligned}\text{Var}(X) &= \sum_{k=1}^{K} (x_k^2 - 2\mu x_k + \mu^2) p_k \\ &= \sum_{k=1}^{K} (x_k^2 p_k - 2\mu x_k p_k + \mu^2 p_k)\end{aligned}$$

Since the $\sum$ (sigma) operation may be distributed across sums or differences of random variables,

$$\text{Var}(X) = \sum_{k=1}^{K} x_k^2 p_k - \sum_{k=1}^{K} 2\mu x_k p_k + \sum_{k=1}^{K} \mu^2 p_k$$

Since

$$E(X) = \sum_{k=1}^{K} x_k p_k$$

one can define in an analogous way

$$E(X^2) = \sum_{k=1}^{K} x_k^2 p_k$$

Thus

$$\begin{aligned}
\operatorname{Var}(X) &= E(X^2) - 2\mu \sum_{k=1}^{K} x_k p_k + \mu^2 \sum_{k=1}^{K} p_k \\
&= E(X^2) - 2\mu E(X) + \mu^2 (1) \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - [E(X)]^2
\end{aligned}$$

This completes the proof.

The use of this computing formula is illustrated for the probability distribution summarized in Table 6-14. For this example,

$$\operatorname{Var}(X) = E(X^2) - [E(X)]^2 = \tfrac{2970}{243} - (\tfrac{810}{243})^2 = 1.111$$

Table 6-14. Example illustrating the computing formula of the variance of a discrete random variable.

x_k	p_k	$x_k p_k$	x_k^2	$x_k^2 p_k$
0	$\frac{1}{243}$	0	0	0
1	$\frac{10}{243}$	$\frac{10}{243}$	1	$\frac{10}{243}$
2	$\frac{40}{243}$	$\frac{80}{243}$	4	$\frac{160}{243}$
3	$\frac{80}{243}$	$\frac{240}{243}$	9	$\frac{720}{243}$
4	$\frac{80}{243}$	$\frac{320}{243}$	16	$\frac{1280}{243}$
5	$\frac{32}{243}$	$\frac{160}{243}$	25	$\frac{800}{243}$
Total		$\frac{810}{243}$		$\frac{2970}{243}$

This example should be immediately recognized as the probability distribution of a binomial random variable with $N = 5$ trials and the probability of success given by $p = \frac{1}{3}$. In this case, the expected value of this binomial is given by $E(X) = Np = (5)(\frac{1}{3}) = 1.6667$ and its variance is given by $\operatorname{Var}(X) = Npq = (5)(\frac{1}{3})(\frac{2}{3}) = \frac{10}{9} = 1.1111$.

6-14 PARAMETERS OF PROBABILITY DISTRIBUTIONS

The quantities N and p associated with a binomial random variable are members of a special class of functions that statisticians call *parameters*. As soon as these particular numbers are known, any probability statement can be made about the variable since these numbers contain all of the information about the distribution. For example, if $N = 64$ and $p = .8$, then

$$P(X = 20) = \binom{64}{20}(.8)^{20}(.2)^{44}$$

$$P(X \geqslant 60) = \sum_{x=60}^{64} \binom{64}{x}(.8)^{x}(.2)^{64-x}$$

$$P[(X_1 = 18) \cup (X_2 = 40)] = \binom{64}{18}(.8)^{18}(.2)^{46} + \binom{64}{40}(.8)^{40}(.2)^{24}$$

$$E(X) = Np = 64(.8) = 51.2$$

$$\text{Var}(X) = Npq = 64(.8)(.2) = 10.24$$

$$\sigma_X = \sqrt{10.24} = 3.2$$

The important thing to note about parameters is their close connection to a probability distribution. Experiments are conducted and sample surveys are performed mainly with the purpose of determining the numerical values to assign to parameters of known probability distributions. In fact, the most interesting problems of statistics are directly related to questions about parameters. These questions are:

1. How can one best estimate parameters?
2. How can one make inferences about probability distributions on the basis of estimated parameters?

Much of the sequel will be concerned with these two questions.

Binomial random variables with parameters N and p will be denoted in the remaining chapters as $B(N,p)$. This symbol should be interpreted as: X is binomial with parameters N and p. In like manner, the uniform distribution that ranges over the values $a, a+1, a+2, \ldots, b$ will be denoted in the sequel by $U(a,b)$ While it has not been explicitly shown, $E(X) = (a+b)/2$ and $\text{Var}(X) = (b^2 - a^2)/12$. For the dice model, one has $E(X) = (1+6)/2 = \frac{7}{2} = 3.5$ and $\text{Var}(X) = (6^2 - 1^2)/12 = \frac{35}{12} = 2.9167$, as were previously obtained.

6-15 SUMMARY

Discrete random variables represent the most frequently encountered criterion variable of behavioral research. In general, a discrete variable X is said to be a

discrete random variable if for each possible outcome one can associate a probability of occurrence such that the sum of the probabilities is unity.

One of the important properties of simple random variables is that one can create complex random variables with greater intrinsic value. In the case of repeated measurement on the variable X, one can generate a new variable $T = X_1 + X_2 + \cdots + X_N$, which represents the total outcome on N trials, If, for example, X refers to the family income of boys brought before juvenile court in a particular city, then T represents the total family income of N boys brought before the court for sentencing.

As a discussion aid, discrete probability distributions can be illustrated by drawing a line graph or cumulative probability polygon, more commonly called a step graph. From the line graph, it is easy to see whether a discrete probability distribution is symmetric or skewed. Since graphs can be easily manipulated to deceive a reader, they are not customarily used for analytical investigations. For the most part, their utility is restricted to descriptive presentations. For analytical procedures, mathematically defined summary measures find greater use. For example, the mode of a probability distribution is defined as the value of X that has maximum probability, provided that such a value exists. The median of a probability distribution is defined as the value of X that partitions the probability so that if $X = M$ is the median, $P(X \leqslant M) = P(X \geqslant M) = \frac{1}{2}$. Some discrete probability distributions have an infinite number of medians existing between two numbers, M_1 and M_2. When this occurs, the median is defined as $M = \frac{1}{2}(M_1 + M_2)$.

The most important measure of central tendency is the expected value or weighted average of a probability distribution, where the weights are the probabilities of occurrence. Mathematically,

$$E(X) = \sum_{k=1}^{K} x_k P(X = x_k)$$

Operationally, the expected value is the long-term average of the distribution. If, for example, it is found that over a long period of time the average family income of boys brought before juvenile court is \$7,283.68, one would treat this number as though it were the expected family income of boys brought before juvenile court. The rationale for this behavior can be traced to the relative-frequency interpretation of probability.

An important property of expected values is stated in Theorem 6-1. This theorem states that if $T = X_1 + X_2 + \cdots + X_N$, then $E(T) = \sum_{k=1}^{N} E(X_k)$. If the X's have identical distribution, then $E(T) = NE(X)$. Thus, if eight boys are brought into juvenile court on a particular day, the expected total family income of these eight boys should be close to $E(T) = 8(\$7,283.68) = \$58,269.44$.

As a special case of this last result, it was seen that if T is $B(N,p)$, then $E(T) = Np$. Thus, if over many years of experience it is found that 16 percent of the boys brought before juvenile court come from homes in which a father is not present, then in the

set of eight boys brought before juvenile court the expected number of boys without a father at home is given by $E(T) = 8(.16) = 1.28$.

Probability distributions differ with respect to spread. Some extend over a large range of values while others tend to be compact and dense. The variance and standard deviation of the probability distribution describe its spread. Distributions with large variability have large variances and standard deviations. Distributions that extend over a small range of values have small variances and standard deviations. Like the mode, median, and expected value, these measures are defined as mathematical properties of a probability distribution. For the variance, one has

$$\sigma_X^2 = \mathrm{Var}(X) = \sum_{k=1}^{K} (x_k - \mu)^2 P(X = x_k)$$

To determine the standard deviation, one need only determine the square root of the variance, that is, $\sigma_X = \sqrt{\mathrm{Var}(X)}$. One of the major uses of variances is to compare the variability that exists between different populations. For example, it might be found over many years of experience that the standard deviation in family income of boys brought to juvenile court varies with the age of boys, so that for boys under 10, $\sigma_1 = \$1{,}263.17$, while for boys aged 10 to 18, $\sigma_2 = \$823.16$. Clearly, there is greater variance in family income for delinquent juvenile boys under 10 years of age than there is for boys over 10 years of age. The relative differences are given by

$$E = \frac{\sigma_2^2}{\sigma_1^2} = \frac{(823.16)^2}{(1{,}236.17)^2}(100) = 44.4\%$$

Thus, the variance in family income for the older boys is only 44.4 percent that of the younger boys.

The important property of variances is stated in Theorem 6-4. If $T = X_1 + X_2 + \cdots + X_N$, then $\mathrm{Var}(T) = \sum_{k=1}^{N} \mathrm{Var}(X_k)$, provided that the variables are independent. As a special case, it was noted that when the variances are equal, $\mathrm{Var}(T) = N\sigma^2$. If five of the eight boys brought before the juvenile court are less than 10 years old, then the variances in family income for the families is given by $5(1{,}236.17)^2$, so that $\sigma_T = \$2{,}764.08$.

As another special case, it was noted that for a binomially distributed variable, $\mathrm{Var}(T) = Npq$. Thus, for the set of eight juveniles, the variance for the number without a father at home is given by $\mathrm{Var}(T) = 8(.16)(.84) = 1.0752$ with $\sigma = \sqrt{1.0752} = 1.03$.

EXERCISES

***6-1.** In a large urban secondary school, students were asked how many movies they attended during the previous month. The results were as shown in the following table.

Number of movies	$P(X=x)$
0	20
1	.05
2	.03
3	.08
4	.16
5	.27
6	.09
7	05
8	.02
9	02
10	01
11	01
12	.01

Draw the line graph representing this distribution. Is the distribution symmetrical? Explain. Draw the corresponding step diagram. From the graph, determine the value of X that represents the median Also, determine the value of X that divides the population so that 80 percent of the students are below X.

***6-2.** Determine the mode, median, and expected value of the distribution of Exercise 6-1. What do these figures say about the distribution?

***6-3.** Compute the variance and standard deviation of the distribution of Exercise 6-1 In the previous school year the same question was asked about movie attendance for the same month. At that testing it was found that $\sigma = 1.31$ What does this suggest about movie attendance on the second testing as compared to the first testing?

***6-4.** If five students were selected from the population of Exercise 6-1 and if $T = X_1 + X_2 + X_3 + X_4 + X_5$ gives the total number of movies attended, what are $E(T)$ and σ_T? What interpretation can you give to these numbers?

***6-5.** A teacher plans to give a spelling test consisting of four words that students have a difficult time spelling correctly. Over four years of teaching she has determined the probabilities of the correct and incorrect spelling of each word. These probabilities are given by $\{p_1 = .4, q_1 = .6\}$, $\{p_2 = .3, q_2 = .7\}$, $\{p_3 = .3, q_3 = 7\}$, and $\{p_4 = .3, q_4 = .7\}$. Let X = number of correctly spelled words. Find the probability distribution of X, its expected value, and the standard deviation. Note that if the probabilities of misspelling are independent, $P(X=4)$ is given by $P(X=4) = p_1p_2p_3p_4 = (.4)(.3)(.3)(.3) = .0108$, while $P(X=0) = q_1q_2q_3q_4 = (.6)(.7)(7)(.7) = .2058$. When the test was given to 25 children, 6 of them obtained perfect scores. Should the teacher be surprised? Explain.

6-6. (*a*) Find the expected value and variance of the probability distribution of Exercises 4-3 4-7 and 4-8

(*b*) On the sample space of Figure 6-2, define the following random variable

$$S^2 = \left(X_1 - \frac{X_1 + X_2}{2}\right)^2 + \left(X_2 - \frac{X_1 + X_2}{2}\right)^2$$

For (1,1), $S^2 = 0$, while for (1,2), $S^2 = (1 - \frac{3}{2})^2 + (2 - \frac{3}{2})^2 = (-\frac{1}{2})^2 - (\frac{1}{2})^2 = .50$. Determine the probability distribution of S^2 and find the expected value and variance. The distribution generated by this variable is called the sampling distribution of S^2 for samples of size 2 selected from the uniform distribution $U(1,6)$. Distributions of this nature are studied in Chapter 10

***6-7.** In a certain city it is known that 30 percent of youths sentenced for stealing cars return to court on a second charge for the same crime. If, during the weeks of a particular month, 16, 24, 33, and 27 youths are convicted of car stealing, determine the expected number of returnees among those sentenced each week of the month. Also determine the standard deviation of returnees for each week and for the entire month. If only 10 of these youths come back for car stealing, what would this indicate?

***6-8.** Determine the $E(X)$ and σ_X for the probability distributions shown in Tables 5-9 and 5-10. What does this tell you about the variability in X when sampling with or without replacement?

***6-9.** In attitude studies, it is common practice to employ questionnaires and items whose response choices make up an ordered scale but for which the response choices are qualitative. An example of such a response set is {strongly disagree, moderately disagree, moderately agree, strongly agree}. One way to analyze responses to these kinds of items is to superimpose a Likert scale onto the response set. For this, one associates a score of +1 for strongly disagree, a score of +2 for moderately disagree, a +3 for moderately agree, and a +4 for strongly agree. *Assume* that the relative frequencies of the margins in the table of Exercise 3-10 represent probabilities of the response set. Determine the expected values and standard deviations of each distribution. What do these measures indicate about an individual's attitudes toward the two items of the questionnaire?

***6-10.** In the analysis of the data of Table 3-2, it was found that attitude toward school was related to the school the student attended. Suppose 25 students had been selected from each of the two schools. What would be the expected number of students who select "more" as a response choice at the two schools?

7 CONTINUOUS PROBABILITY DISTRIBUTIONS

Although the uptrend in adult size accounts for a part of the increase we find in children, it is clear that a far greater part of the increase is related to the earlier age at which children mature. The trend toward earlier maturation is perhaps best shown by statistics on the age of menarche, or first menstrual period. (The age of the first appearance of pubic hair in boys is a less reliable index of puberty) It is known from studies in which individuals were followed through childhood until menarche that, in a given population, age at menarche is distributed in a bell-shaped curve called the normal distribution This makes it possible to estimate the mean age of menarche on a "cross-sectional" basis What one does is select a proper sample—for instance a sample of all the schools in a certain area—and then ask every girl in the sample group whether or not she has experienced her first menstrual period. Ideally all girls between the ages of nine and 17 should be interrogated Most large-scale modern studies are carried out in this way An equally valid procedure is the "longitudinal" study, in which every child is checked repeatedly until menarche

.. The main conclusion is perfectly clear girls have experienced menarche progressively earlier during the past 100 years by between three and four months per decade. On this basis puberty is attained $2\frac{1}{2}$ to $3\frac{1}{2}$ years earlier today than it was a century ago. The trend in height and weight at the age of puberty is in good agreement with this figure, the 11-year-old children of today having the size of 12-year-olds 30 or 40 years ago.

7-1 CONTINUOUS RANDOM VARIABLES

While qualitative and discrete random variables are encountered in great abundance in behavioral research, continuous random variables, nevertheless, play the prominent role in the analysis of behavioral data, not because behavioral data tends to be continuous, but because the mathematics and distribution theory of continuous variables is simpler and serves as a good approximation for the probability distributions of discrete random variables. While it cannot be stated with certainty, it is perhaps true that binomial and related random variables are the most frequently employed variables of behavioral research. Unfortunately, when N, one of the parameters of a binomial variable, is large, probability computations become time-consuming, involved, and difficult to carry out. Fortunately, a naturally defined mathematically continuous variable can be used to practical advantage to provide adequate approximations to the correct binomial probabilities. Of even greater importance is that this same mathematically defined continuous variable can serve as an excellent approximation for the probability distributions of such diverse discrete variables as number of correct choices a rat makes in running a complex maze or IQ measures of first-grade children on the Stanford-Binet intelligence test. Furthermore, many other achievement, attitude, or aptitude tests that are in common use among behavioral researchers can be succinctly described by this same continuous probability distribution. For this reason, the algebra and descriptive properties of continuous random variables will be studied in considerable detail. Particular attention will be focused on the properties of the continuous distribution called the normal distribution.

As stated in Chapter 1, a variable X is said to be continuous over the range of values $a \leqslant X \leqslant b$ if it can assume all possible numerical values in this range. Variables that are of this nature are weight, length, and height. Other examples are velocities, masses, pressures, times, volumes, etc. The total set of numerical values that these variables can assume in any one context cannot be enumerated; i.e., it is impossible to count or list the totality of outcomes. This inability to list the complete set of outcomes can be illustrated by the variable of height measured on all human beings, past, present, or future, that reach a twenty-first birthday. If it were possible to measure the heights of everyone on their twenty-first birthday, it would be found that every number between two limits, say 24 inches and 90 inches, would occur. This means that all numbers between these two limits are possible values to indicate height. As a result, there will exist a person whose height at age twenty-one is equal to 69.36243 inches, while someone else will be 69.36244 inches tall. Furthermore, there is bound to be someone whose height is between these numbers. For emphasis it is recalled that measurement refinement of this order is not possible for discrete variables such as scores on a true-false test that contains 50 items. While the possible range of correct scores is $0 \leqslant X \leqslant 50$, the complete set of possible scores consists of the countable set of 51 unique values: $\{0,1,2,3,\ldots,50\}$. A score such as 23.78945 is impossible.

For a discrete variable in a finite sample space, one can always determine

$$P(X=x)=\frac{n(X=x)}{N(S)}=p_x$$

If the universe is infinite, this probability can be estimated by direct application of the relative-frequency definition of probability. For a continuous variable, the probability that X can assume a specific value is 0. Thus, as will be shown, $P(X=x)=0$ if X is a continuous random variable. Since it is impossible to specify the probability of any unique outcome, one cannot define a continuous random variable in a manner similar to that used for discrete random variables. As a result, one must resort to some other device for specifying the probabilities associated with a continuous variable. This definition is achieved by specifying the probability that a continuous variable X lies in a certain interval of values. In the most general case, the range of X extends from minus infinity to plus infinity: $\{-\infty < X < \infty\}$. Without much difficulty, this infinite interval can always be partitioned into a finite set of *mutually exclusive and exhaustive intervals*, as is shown graphically in Figure 7-1.

On the basis of this partition, the definition of a continuous random variable is easily made. A continuous variable is said to be a continuous random variable if it is possible to assign a probability distribution to its range of values. For the partitioned set of Figure 7-1, this could be done as follows:

$$\begin{aligned}
P(-\infty < X < a_1) &= p_1\\
P(a_1 \leqslant X < a_2) &= p_2\\
&\vdots\\
P(a_{K-1} \leqslant X \leqslant +\infty) &= p_K
\end{aligned}$$

provided that

$$p_1+p_2+\cdots+p_K=1$$

7-2 GRAPHIC METHODS FOR CONTINUOUS VARIABLES

When a continuous probability distribution is defined in terms of probabilities assigned to mutually exclusive and exhaustive subsets, one can graphically portray the distribution by drawing a graph called a *histogram*. To see how this graph is constructed, suppose a psychologist is conducting a study in which the dependent variable of interest is the number of seconds it takes college sophomores to solve a problem involving the discrimination of two colors of nearly equal intensity. Suppose the probability distribution of the times is as shown in Table 7-1. A graphic

Figure 7-1. Partitioning of the continuous interval $-\infty < X < \infty$ into K mutually exclusive and exhaustive subsets.

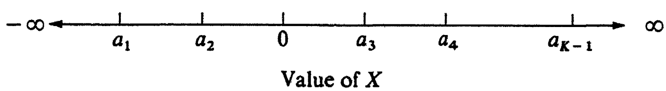


Table 7-1. Probability distribution of time it takes to discriminate between two colors of nearly equal intensity.

Time interval in seconds	*Probability that time to solve problem is in interval*	*Cumulative probabilities*
$0 \leqslant X < 5$	.05	.05
$5 \leqslant X < 10$	.23	.28
$10 \leqslant X < 15$	.39	.67
$15 \leqslant X < 20$	.17	84
$20 \leqslant X < 25$	.10	94
$25 \leqslant X < 30$	.04	98
$30 \leqslant X < 35$	.01	99
$35 \leqslant X < 40$	01	1 00
Total	1.00	

representation of this distribution is shown in Figure 7-2. In constructing this graph, draw bars over the range of the intervals equal in height to the probabilities. Thus, over the interval $0 \leqslant X < 5$, a bar is constructed with height equal to .05. This procedure is then repeated for each interval. As Figure 7-2 suggests, the probability distribution of times is positively skewed. Thus, just like discrete random variables, continuous random variables can have symmetric, positively skewed, or negatively skewed distributions.

Another way to graphically represent continuous distributions defined over intervals is to construct a *probability polygon.* To construct this graph, determine the midpoints of each interval and instead of drawing bars equal in height to the probability, join the midpoints of the bars to make a closed geometric polygon. This is illustrated in Figure 7-3.

Finally, another and quite useful graph is available in the *cumulative probability polygon.* For this graph, determine the cumulative probabilities $\{P(X \leqslant a_1), P(X \leqslant a_2), \ldots, P(X \leqslant \infty)\}$ and plot these probabilities against the various values of $\{a_1, a_2, \ldots, a_{k-1}, \infty\}$. The cumulative probability polygon for the distribution of Table 7-1 is shown in Figure 7-4. From this graph, the median time is seen to be 13 seconds. This value is found by drawing a line horizontal to the X axis, passing through $P(X \leqslant a) = .50$ to the probability polygon. At the point the line cuts the graph, a perpendicular is dropped to the X axis. The value of X where the perpendicular cuts the X axis is the median. In this case, $M = 12.8$.

7-3 QUARTILES, DECILES, AND PERCENTILE VALUES OF PROBABILITY DISTRIBUTIONS

Not only may the median value of X be determined from a cumulative probability polygon, but any other percentage value may also be determined. For example,

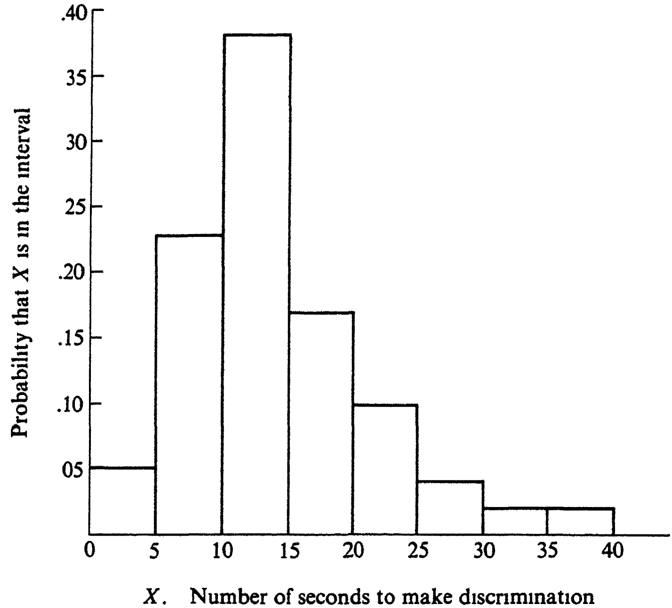


Figure 7-2. Histogram of the hypothetical probability distribution of time it takes in seconds to discriminate between two colors of nearly equal intensity.

the first quartile, Q_1, the value of X that divides the distribution such that $P(X \leqslant Q_1) = .25$, can be found by drawing a line from the 25-percent point on the vertical scale over the graph and then dropping a perpendicular to the horizontal axis. Where the perpendicular cuts the horizontal axis, the value of Q_1 is defined.

The second quartile, Q_2, is equal to the median. The third quartile, Q_3, is the value of X that satisfies the probability equation $P(X \leqslant Q_3) = .75$. The value of X that cuts off the lower 10 percent of the distribution is called the first decile or the 10th percentile. Clearly, the first quartile is synonymous with the 25th percentile. The median is identical to the 50th percentile, while the 75th percentile is equal to the third quartile. The values of X corresponding to the quartiles, deciles, and percentiles can be determined with ease on a cumulative probability polygon by following the procedure described for the median and the first quartile. The determination of P_{85}, the 85th percentile of the distribution of times to solve the problem, is illustrated in Figure 7-4. In this case, $P_{85} = 20.5$ seconds.

7-4 CONTINUOUS DISTRIBUTIONS DEFINED AS SMOOTH CURVES

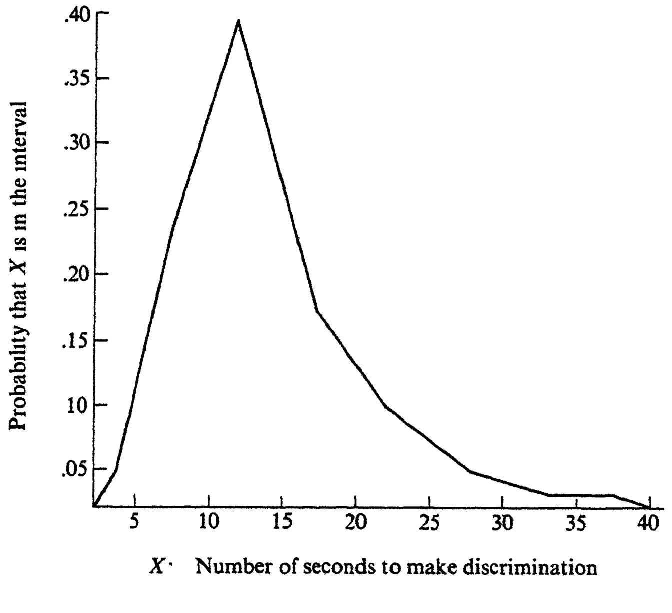
In most applications, the range of X is not partitioned into a finite set of mutually exclusive and exhaustive intervals. Instead, a particular form for the probability

distribution is assumed or stated by a mathematical equation and then the probabilities of any intervals are determined by advanced mathematical methods. Continuous probability distributions defined in this manner can be graphically portrayed, as shown in Figure 7-5. As these four examples show, there are many different forms possible for continuous probability distributions.

Drawing 1 is an example of a continuous variable with a positively skewed distribution over the range of $0 \leqslant X < \infty$. Drawing 2 represents a symmetric distribution balanced at $X = 0$. A bimodal distribution is represented by drawing 3. Distributions of this type generally indicate that two overlapping distributions exist. A careful researcher would try to identify the sources of the two peaks and then try to separate the distributions. Distributions such as drawing 4 are unusual in behavioral research. Distributions of this form are said to be J shaped over the finite range $a \leqslant X \leqslant b$.

With this type of graphic representation, it is easy to show why the probability is 0 that a continuous variable takes on any numerical value. To see this, consider the interval $x \pm \Delta$ drawn about the point x in Figure 7-6. Since area corresponds to probability, the area under the curve is 1 and the probability that x is between $x - \Delta$ and $x + \Delta$ is given by the shaded area. Suppose that Δ is reduced. When this occurs, the probability of the shaded area gets smaller. In the limit when $\Delta = 0$, the $P(X = x)$ must be 0, since the shaded area must reduce to 0.

Figure 7-3. Probability polygon for the hypothetical distribution of time it takes in seconds to discriminate between two colors of nearly equal intensity.



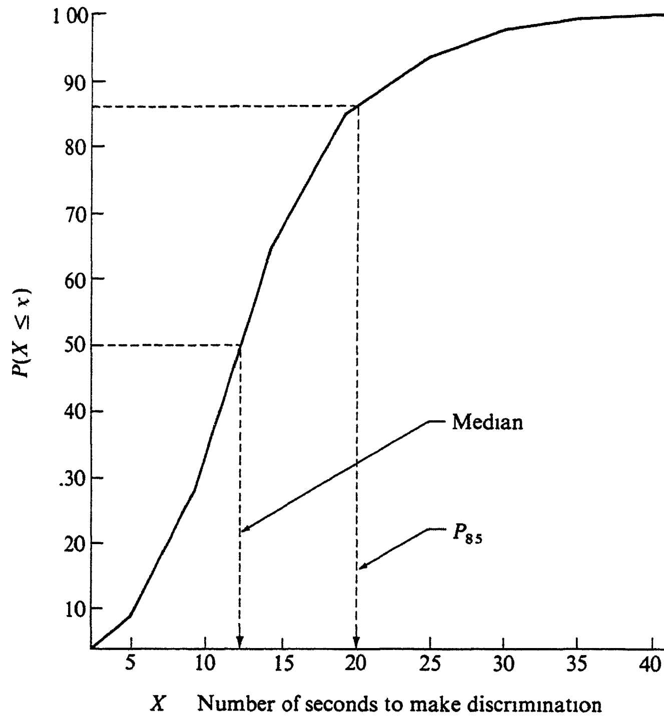


Figure 7-4. Cumulative probability polygon for the hypothetical probability distribution of time it takes in seconds to discriminate between two colors of nearly equal intensity.

7-5 THE NORMAL DISTRIBUTION

One particular continuous distribution that has found widespread use by statisticians and researchers is a distribution called the *normal* distribution. This distribution is not "normal" in the usual dictionary meaning of the word. For some unknown reason this important distribution has been misnamed. As a result, a normal distribution is not to be contrasted with an abnormal distribution; that is not the significance of the term "normal." While it is true that many variables in nature seem to obey probability laws that are adequately described by a normal probability model, this distribution is of greater importance because it has some simple mathematical properties that are useful for behavioral research studies. Some statisticians always refer to this distribution as the gaussian distribution in honor of Gauss (1777–1855), who studied its probability properties in great detail.

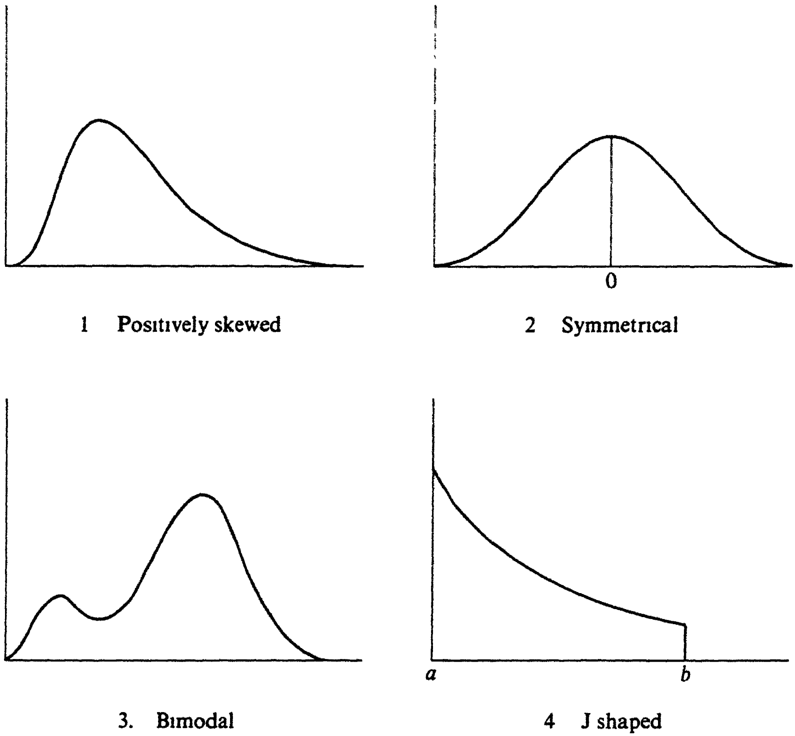
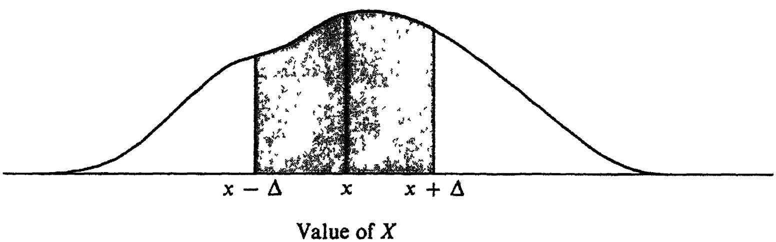


Figure 7-5. Four examples of continuous probability distributions

Figure 7-6. Demonstration that $P(X = x) = 0$ for continuous random variables



The normal distribution is a member of the larger set of unimodal symmetrical distributions. This means that its expected value $E(X)$, its median M, and its mode $\mathscr{M}$ are all equal to the same value. Therefore

$$P[X > E(X)] = P(X > \mathscr{M}) = P(X > M) = .50$$

Furthermore,

$$P[X = E(X)] = P(X = \mathscr{M}) = P(X = M) = 0$$

The general form of the distribution is as shown in Figure 7-7. The center of the distribution is at the expected value. The curve has two changes of direction. Coming from the left, the curve is concave up. It increases with an increasing slope, and then begins to increase at a decreasing rate, until it levels off at the expected value. It then decreases, with an increasing rate, and then again it becomes concave up and decreases with a decreasing rate. The two points at which the curve changes its concavity are called its inflection points. These inflection points occur at one standard deviation above the expected value and one standard deviation below the expected value. Thus

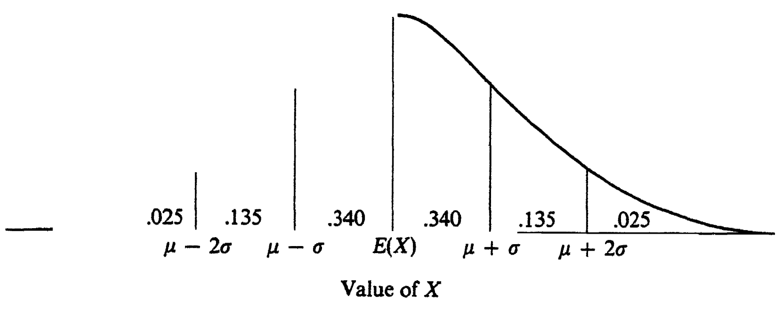
$$\text{Lower inflection point} = E(X) - \sqrt{\text{Var}(X)} = \mu - \sigma$$
$$\text{Upper inflection point} = E(X) + \sqrt{\text{Var}(X)} = \mu + \sigma$$

These inflection points have the following properties:

$$P(X < \mu - \sigma) = .16 = P(X > \mu + \sigma)$$
$$P(\mu - \sigma < X < \mu) = .34 = P(\mu < X < \mu + \sigma)$$

As one might suspect, the probabilities under the normal curve can be determined as soon as the expected value and the standard deviation are known. These are the parameters of the normal distribution.

Figure 7-7. A typical normal distribution where probabilities are reported for intervals that are one standard deviation wide



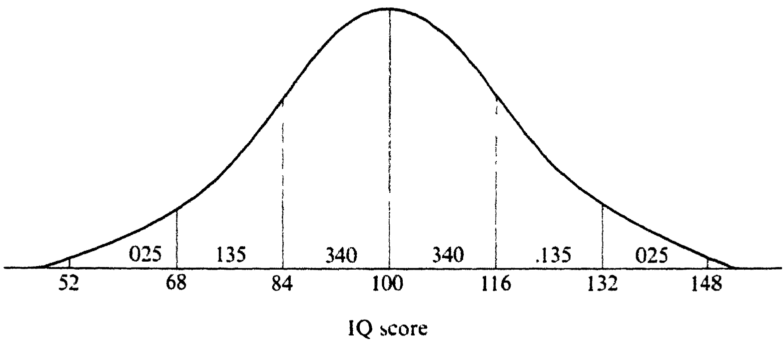


Figure 7-8. Distribution of IQ scores with the IQ scale in standard deviation units

Some properties of the normal distribution that are stated without proof but that should be committed to memory are the following:

1. $P(\mu - \sigma < X < \mu + \sigma) = .68$
2. $P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = .95$

2*a*. $P(\mu - 2\sigma < X < \mu + 2\sigma) = .954$

3. $P(\mu - 2.58\sigma < X < \mu + 2.58\sigma) = .99$
4. $P(\mu - 3\sigma < X < \mu + 3\sigma) = .9997$

Statements 2 and 2*a* state that about 95 percent of the total probability is between plus or minus two standard deviations. Statement 4 indicates that for all practical purposes, 100 percent of the distribution is between plus and minus three standard deviations.

7-6 EXAMPLE OF A NORMALLY DISTRIBUTED RANDOM VARIABLE

Most psychologists assume that intelligence has a continuous distribution within the population. Unfortunately, this variable is extremely difficult to measure and so approximate methods have been developed to give some indication of a person's native intelligence. However, it should be noted that these measures are truly discrete even though the underlying assumption is made that intelligence is distributed continuously in the population.

There are many tests made by educators and psychologists to measure intelligence. One well-known test has been artificially constructed to approximate a normal distribution with an expected IQ score of 100 and a standard deviation of 16 units. This distribution of IQ scores can be represented as shown in Figure 7-8. Even though the distribution of IQ scores is discrete, let it be assumed in the following examples that it is continuous. This means that the correction for continuity can be ignored. This will be done only to simplify the discussion of this section. In Section 7-9, the appropriate correction for continuity will be considered for this particular vari-

able. If a person is to be selected at random from this distribution of IQ scores, the probability that his IQ will be above 100 is given by:

1. $P(X > 100) = P(X > \mu) = .50$

The probability that his IQ will be between 100 and 116 is given by:

2. $P(100 < X < 116) = P(\mu < X < \mu + \sigma) = .34$

The probability that his IQ will be between 84 and 116 is given by:

3. $P(84 < X < 116) = P(\mu - \sigma < X < \mu + \sigma) = .68$

The probability that his IQ will be above 116 or less than 68 is given by:

4. $$\begin{aligned} P[(X > 116) \cup (X < 68)] &= P(X > 116) + P(X < 68) \\ &= P(X > \mu + \sigma) + P(X < \mu - 2\sigma) \\ &= .160 + .025 \\ &= .185 \end{aligned}$$

The probability that his IQ will be greater than 104 satisfies the inequality:

5. $P(X > 116) < P(X > 104) < P(X > 100)$

Clearly,

$$.16 < P(X > 104) < .50$$

Up to this point, the information presented concerning the probabilities of normal variables does not permit the exact determination of this probability. Fortunately, this probability can be computed exactly since the normal distribution has been tabled. The use of these tables will be illustrated following the algebra of expected values.

7-7 THE ALGEBRA OF EXPECTED VALUES

Four important theorems are stated and proved for discrete random variables. Comparable proofs for continuous random variables are beyond this book. However, the theorems are identical for discrete and continuous variables. While the proofs are for discrete variables, the examples are based on continuous variables.

Theorem 7-1

If the same constant A is added to all possible values that a random variable X can assume, then the expected value of the transformed variable Y is the same as the expected value of X plus the value of the constant A. Thus, if $Y = X + A$, then $E(Y) = E(X) + A$.

Proof. Since $y_k = x_k + A$, then

$$P(Y = y_k) = P(X = x_k) = p_k$$

By definition,

$$\begin{aligned} E(Y) &= \sum_{k=1}^{K} y_k p_k \\ &= \sum_{k=1}^{K} (x_k + A) p_k \\ &= \sum_{k=1}^{K} (x_k p_k + A p_k) \\ &= \sum_{k=1}^{K} x_k p_k + \sum_{k=1}^{K} A p_k \\ &= E(X) + A \sum_{k=1}^{K} p_k \\ &= E(X) + A \end{aligned}$$

This completes the proof.

As an example, assume that heights of men are normally distributed with $E(X) = 68$. If 5 inches were added to everyone's height, then the transformed height distribution has

$$E(Y) = E(X + 5) = E(X) + 5 = 68 + 5 = 73$$

Theorem 7-2

If the same constant A is added to all possible values that a random variable X can assume, then the variance of the transformed variable Y is the same as the variance of X. Thus, if $Y = X + A$,

$$\mathrm{Var}(Y) = \mathrm{Var}(X + A) = \mathrm{Var}(X)$$

Proof. By definition,

$$\begin{aligned} \mathrm{Var}(Y) &= \sum_{k=1}^{K} [y_k - E(Y)]^2 p_k \\ &= \sum_{k=1}^{K} [(x_k + A) - (\mu_x + A)]^2 p_k \\ &= \sum_{k=1}^{K} (x_k - \mu_x)^2 p_k \\ &= \mathrm{Var}(X) \end{aligned}$$

This completes the proof.

Thus, the addition or subtraction of a constant does not affect the variance of a random variable, but it does affect the expected value. This should be intuitively obvious. As an example, consider all men whose height is 70 inches. The equation $Y = X + 5$ transforms all of them to the new height value of 75. All persons with a height of 69 inches are transformed to the new height value of 74. As this suggests, everyone's relative position to one another remains the same. Individuals who differed by 1 inch on the X scale continue to differ from one another by 1 inch on the Y scale. This indicates that the form of the distribution is unaffected by the addition or subtraction of a constant numerical value. Thus, if the distribution of heights has $\sigma_X = 4$ inches, then $\sigma_Y = 4$ inches. These properties are illustrated in Figure 7-9.

Theorem 7-3

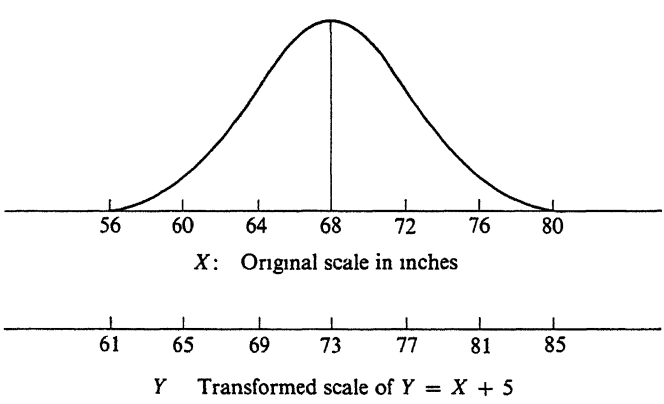
If the same constant B is used to multiply all possible values that a random variable X can assume, then the expected value of the transformed variable W is also multiplied by that constant B. Thus, if $W = BX$, then

$$E(W) = BE(X)$$

Proof Since $w_k = Bx_k$,

$$P(W = w_k) = P(X = x_k) = p_k$$

Figure 7-9. The relationship between the distribution of male heights in inches and the transformed distribution of heights in which each height is increased by 5 inches.



By definition,

$$
\begin{aligned}
E(W) &= \sum_{k=1}^{K} w_k p_k \\
&= \sum_{k=1}^{K} (Bx_k) p_k \\
&= B \sum_{k=1}^{K} x_k p_k \\
&= BE(X)
\end{aligned}
$$

This completes the proof.

As an example, suppose the population of males had their heights measured with a yardstick that was marked in inches, and suppose it were found that $E(X) = 68$ inches. If the measurements were desired in feet, one would compute $W = \frac{1}{12}X$. Thus

$$
\begin{aligned}
E(W) = E(\tfrac{1}{12}X) &= \tfrac{1}{12}E(X) \\
&= \tfrac{1}{12}(68) \\
&= 5 \text{ feet, } 8 \text{ inches} \\
&= 5\tfrac{2}{3} \text{ feet}
\end{aligned}
$$

Whereas the addition or subtraction of a constant does not affect the shape of a probability distribution, multiplication by a constant changes both the scale and the shape of the distribution. This is expressed in the next theorem.

Theorem 7-4

If the same constant B is used to multiply all possible values that a random variable X can assume, then the variance of the transformed variable W is equal to the variance of X times the square of B. Thus, if $W = BX$, then $\text{Var}(W) = B^2\,\text{Var}(X)$.

Proof. By definition,

$$
\begin{aligned}
\text{Var}(W) &= \sum_{k=1}^{K} [w_k - E(W)]^2 p_k \\
&= \sum_{k=1}^{K} (Bx_k - B\mu_x)^2 p_k \\
&= \sum_{k=1}^{K} B^2(x_k - \mu_x)^2 p_k \\
&= B^2\,\text{Var}(X)
\end{aligned}
$$

This completes the proof.

As an example, suppose the standard deviation in heights on the inch scale for the men of the previous example was 4 inches. The variance on the foot scale is given as

$$\mathrm{Var}(W) = B^2\,\mathrm{Var}(X) = (\tfrac{1}{12})^2(16) = \tfrac{1}{9} = \sigma_W^2$$

so that the standard deviation is given by $\sigma_W = \frac{1}{3}$ foot. This is reasonable since 4 inches are equal to $\frac{1}{3}$ foot. This is illustrated graphically in Figure 7-10.

With these theorems it is possible to transform any random variable to any center and to any scale desired. In particular, there are some special transformations that are used extensively in research. One of the most useful is described in the following two theorems.

Theorem 7-5

If X is transformed by the equation

$$Z = \frac{X - E(X)}{\sigma_X}$$

then

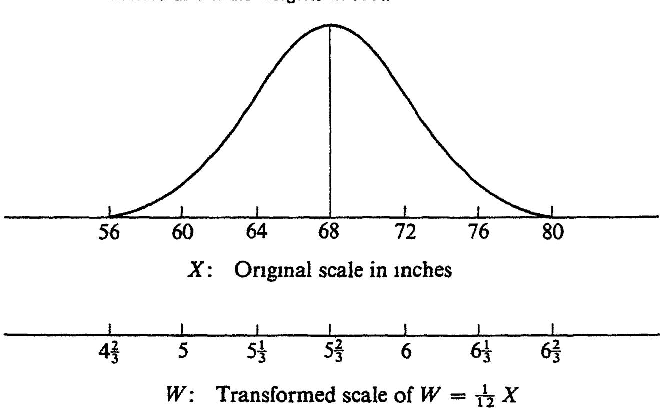
$$E(Z) = 0$$

Proof. By definition,

$$E(Z) = E\left(\frac{X - E(X)}{\sigma_X}\right)$$

$$= E\left(\frac{X}{\sigma_X} - \frac{\mu_x}{\sigma_X}\right)$$

Figure 7-10. The relationship between the distribution of male heights in inches and male heights in feet.



Since $E(X)$ and σ_X are constants, μ_x/σ_X is also a constant, and therefore, by Theorem 7-2, with $A = -\mu_x/\sigma_X$,

$$E(Z) = E\left(\frac{X}{\sigma_X}\right) - \frac{\mu_x}{\sigma_X}$$

By Theorem 7-3, with $B = 1/\sigma_X$,

$$E(Z) = \frac{1}{\sigma_X} E(X) - \frac{\mu_x}{\sigma_X}$$

$$= \frac{\mu_x}{\sigma_X} - \frac{\mu_x}{\sigma_X} = 0$$

This completes the proof.

Theorem 7-6

If X is transformed by the equation

$$Z = \frac{X - E(X)}{\sigma_X}$$

then

$$\mathrm{Var}(Z) = 1$$

Proof. By definition,

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X - E(X)}{\sigma_X}\right) = \mathrm{Var}\left(\frac{1}{\sigma_X}[X - E(X)]\right)$$

By Theorem 7-4, with $B = 1/\sigma_X$,

$$\mathrm{Var}(Z) = \left(\frac{1}{\sigma_X}\right)^2 \mathrm{Var}[X - E(X)]$$

By Theorem 7-1, with $A = -E(X)$,

$$\mathrm{Var}(Z) = \frac{1}{\sigma_X^2}\mathrm{Var}(X)$$

$$= \left(\frac{1}{\sigma_X^2}\right)(\sigma_X^2)$$

$$= 1$$

This completes the proof.

Theorem 7-7 is an exceedingly important theorem whose proof is beyond this book. In essence, it says that a transformed normal distribution remains normal if a particular transforming equation is employed.

Theorem 7-7

If X is normal with $E(X)$ and $\mathrm{Var}(X)$, and if X is transformed by the variable $Z = [X - E(X)]/\sigma_X$, then Z is normal with $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$.

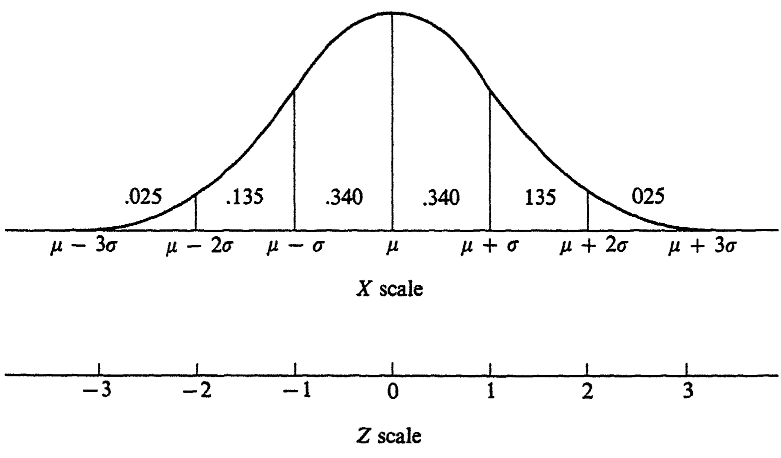
The following example illustrates the meaning of this theorem. Consider the distribution of Stanford-Binet IQ scores with $E(X) = 100$ and $\sigma_X = 16$. The IQ score of 68 is two standard deviations below the expected IQ value of 100, since $68 = 100 - 2(16) = \mu - 2\sigma$. On the Z scale, 68 corresponds to -2 units. As this suggests, the Z scale is in reality a standard deviation scale. A $Z = +1.5$ is a scale value that is 1.5 standard deviations above 0. On the IQ scale it corresponds to $X = \mu + 1.5\sigma = 100 + 1.5(16) = 124$ IQ points. This correspondence between X and Z is shown in Figure 7-11. The variable Z is normal with $E(Z) = 0$, and $\sigma_Z = 1$ is frequently called the standardized normal variable.

7-8 TABLE OF THE NORMAL DISTRIBUTION

The significance of Theorem 7-7 should be well understood. It says that any normally distributed random variable can be transformed into a standard normally distributed random variable with the expectation of 0 and variance of 1. As a result, only one table of the normal distribution is needed. Whenever one wishes to determine the probability of a complex event, one need only transform the variable and look up the probability in the table of the standardized normal variable.

The particular normal distribution that has been tabled has an expected value of 0 and a standard deviation of 1. The table of the normal distribution, Table A-4, included in Appendix A, is a very simple one to read; with a few examples, one should have little difficulty in using it. The numbers appearing in Table A-4 give the cumulative probability distribution for the normal curve with $E(Z) = 0$ and

Figure 7-11. Relationship between X, which is normal with parameters μ_x and σ_X, and $Z = (X - \mu_x)/\sigma_X$, which is normal with $\mu_z = 0$ and $\sigma_z = 1$



$\text{Var}(Z) = 1$. Since the expected value of the normal distribution is 0, it is known that $P(Z < 0) = .5000$. According to the tabled values, $P(Z < 0) = .5000$. As is also known, the 84th percentile of the normal distribution occurs at $Z = 1$. In Table A-4 one reads that $P(Z < 1) = .8413$ As is also known, 68 percent of the distribution lies between -1 and $+1$ standard deviations. Since the table gives cumulative probabilities,

$$\begin{aligned} P(-1 < Z < 1) &= P(Z < 1) - P(Z < -1) \\ &= .8413 - .1587 = .6826 \end{aligned}$$

As this last example shows, the table of the normal distribution is set up to facilitate the determination of the following probability statement: $P(a < X < b) = P(X < b) - P(X < a)$. For example, if $a = -.20$ and $b = .95$, then

$$\begin{aligned} P(-.20 < Z < .95) &= P(Z < .95) - P(Z < -.20) \\ &= .8289 - .4207 = .4082 \end{aligned}$$

7-9 CORRECTION OF DISCRETE SCORES TO FIT A CONTINUOUS PROBABILITY DISTRIBUTION

As mentioned earlier, IQ scores have a discrete probability distribution and since the normal distribution is continuous, some sort of correction should be made to adjust for this discrepancy whenever probabilities of IQ scores are estimated by a normal approximation. Fortunately, the correction is very simple to make and has been introduced in the discussion of discrete random variables. In the literature, this is referred to as correcting for continuity.

As stated earlier, observational data are always measured on a discrete scale, mainly because man is incapable of building measuring devices that are accurate. As a result, measurement tends to be approximate and on a discrete scale. This is true of all measurement and not just the measurement of IQ. To adjust for this condition, consider a person whose IQ is 107. When making the correction for continuity, we treat 107 as representing a score in the continuous range of $106.5 < X < 107.5$. This means that 107 is in an interval of width 1 unit. With this correction, it is now possible to estimate $P(X = 107)$, which would otherwise be set equal to 0. This probability is given by

$$\begin{aligned} P(X = 107) &\cong P(106.5 < X < 107.5) \\ &= P(X < 107.5) - P(X < 106.5) \\ &= P\left(Z < \frac{107.5 - 100}{16}\right) - P\left(Z < \frac{106.5 - 100}{16}\right) \\ &= P(Z < .47) - P(Z < .41) \\ &= .6808 - .6591 = .0217 \end{aligned}$$

In a similar fashion,

$$\begin{aligned} P(X > 104) &= P(X \geqslant 103.5) \\ &= P\left(Z \geqslant \frac{103.5 - 100}{16}\right) \\ &= P(Z \geqslant .22) \\ &= 1 - P(Z \leqslant .22) \\ &= 1 - .5871 = .4129 \end{aligned}$$

7-10 TRANSFORMED Z SCORES AND T SCORES

The four theorems of Section 7-7 are also used extensively by psychometricians and counselors in helping students interpret scores on various scales that measure different characteristics or variables. As an illustration, consider a student who is given a test of manifest anxiety and a test of religious autonomy and suppose he makes the following scores: $\{X_1 = 70$ and $X_2 = 63\}$. As these scores stand, they are uninterpretable, since a score of 70 on the manifest anxiety scale may actually represent a lower score than 63 does on the scale of religious autonomy. Suppose the tests have been standardized so as to be approximately normal with $E(X_1) = 60$ and $\sigma_1 = 12$ and $E(X_2) = 48$ and $\sigma_2 = 8$; then the percentile rankings for these scores are given by

$$\begin{aligned} P(X_1 \leqslant 70) &= P\left(Z \leqslant \frac{70.5 - 60}{12}\right) \\ &= P\left(Z \leqslant \frac{10.5}{12}\right) \\ &= P(Z \leqslant .88) \\ &= .8105 \end{aligned}$$

and

$$\begin{aligned} P(X_2 \leqslant 63) &= P\left(Z \leqslant \frac{63.5 - 48}{8}\right) \\ &= P\left(Z \leqslant \frac{15.5}{8}\right) \\ &= P(Z \leqslant 1.94) \\ &= .9738 \end{aligned}$$

The psychometrician could now tell this student that on the anxiety scale he ranks above 81 percent of his fellow students and that on the autonomy scale he ranks above 97 percent of his fellow students. Thus, even though $X_1 = 70$ is greater than

$X_2 = 63$, X_2 represents a higher score than X_1, since it is further out on the distribution of X_2 scores when measured in standard deviation units.

As stated earlier, the score that divides the class so that 81 percent of the values are below it is called the 81st percentile. In this example, the 81st percentile is equal to 70. In like manner, a score of 63 is said to be the 97th percentile of test 2, since 97 percent of the test scores are equal to or less than 63.

Another way to interpret the student's scores might be to give him a standard score, which has the property that the 50th percentile is at the standard score of 50 and the standard deviation of the scores is 10. The transformation that will do this is given by $T = 10Z + 50$. For this transformation,

$$\begin{aligned} E(T) &= E(10Z + 50) \\ &= 10E(Z) + 50 \\ &= 10(0) + 50 \\ &= 50 \end{aligned}$$

and

$$\begin{aligned} \text{Vat}(T) &= \text{Var}(10Z + 50) \\ &= \text{Var}(10Z) \\ &= 10^2 \,\text{Var}(Z) \\ &= 100(1) \\ &= 100 \end{aligned}$$

While it is necessary to make corrections for continuity when computing percentiles, this procedure is not adhered to when computing standard scores. For the test scores $X_1 = 70$ and $X_2 = 63$,

$$\begin{aligned} T_1 &= 10Z_1 + 50 \\ &= 10\left(\frac{70 - 60}{12}\right) + 50 \\ &= 8.33 + 50 \\ &= 58 \end{aligned}$$

and

$$\begin{aligned} T_2 &= 10Z_2 + 50 \\ &= 10\left(\frac{63 - 48}{8}\right) + 50 \\ &= 18.75 + 50 \\ &= 69 \end{aligned}$$

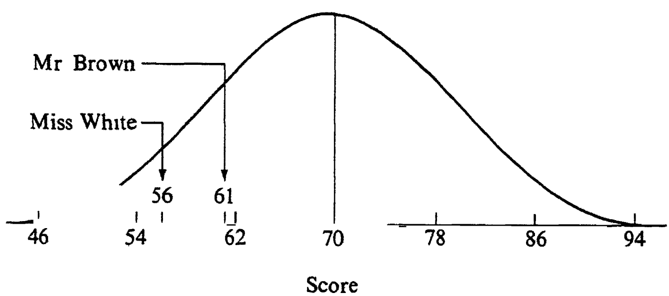
Thus, it is seen again that his percentile ranking for T_2 exceeds his percentile ranking for T_1. Many standardized tests are of this nature. For example, the college board scores are transformed so that $E(T) = 500$ and $\sigma_T = 100$.

T scores are useful for comparing two individuals on the same achievement test or else for comparing one individual's performance on two different achievement tests. As another example, consider administering two standardized tests to two different individuals. Let the expected value for the first test be 70 and the standard deviation for the first test be 8. Let the second test have an expected value of 51 and a standard deviation of 12. Suppose that one of the students, Mr. Brown, obtained a score of 61 on the first test and a score of 40 on the second test, giving him a total score of 101 points. Suppose, also, that Miss White took the same tests and that she had a score of 56 on the first test and a score of 45 on the second test, giving her a total score of 101 points. Even though their scores total the same, their achievement may not be the same. This is because the score of 61 does not have the same significance on the first test as it has on the second test. This is also true for the other possible scores.

If the tests have been standardized so as to give a score distribution resembling a normal distribution with $E(X) = 70$ and $\sigma_X = 8$, then the top score expected would be about $70 + 3(8) = 94$ and the lowest scores expected would be about $70 - 3(8) = 46$. The distribution of scores would be as shown in Figure 7-12. For the second test, the distribution of scores would be as shown in Figure 7-13.

Comparison of these two distributions shows that Mr. Brown and Miss White switched their positions on the second test but both still remained at the lower end of the distribution. Of the two students, one might wish to know which has the lower academic standing. Since the raw scores are not directly comparable, due to the fact that they are in different distributions, it is necessary to convert to the same distribution by means of T scores.

Figure 7-12. Distribution of scores for the first test X_1 along with the scores obtained by Mr. Brown and Miss White



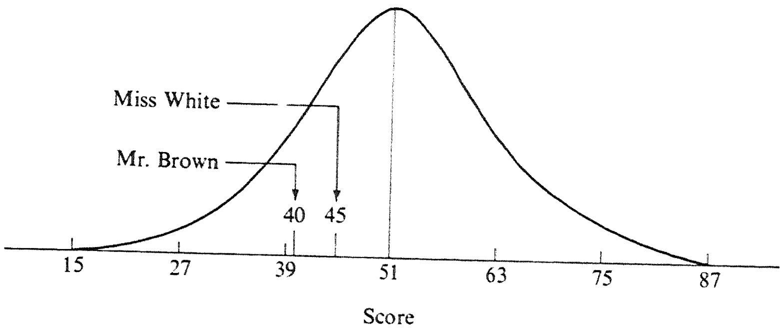


Figure 7-13. Distribution of scores for the second test X_2 along with the scores obtained by Mr. Brown and Miss White.

For Mr. Brown,

$$T_1 = 10\left(\frac{X_1 - E(X_1)}{\sigma_1}\right) + 50$$

$$= 10\left(\frac{61 - 70}{8}\right) + 50$$

$$= -11.25 + 50$$

$$= 39$$

and

$$T_2 = 10\left(\frac{X_2 - E(X_2)}{\sigma_2}\right) + 50$$

$$= 10\left(\frac{40 - 51}{12}\right) + 50$$

$$= -9.17 + 50$$

$$= 41$$

For Miss White,

$$T_1 = 10\left(\frac{X_1 - E(X_1)}{\sigma_1}\right) + 50$$

$$= 10\left(\frac{56 - 70}{8}\right) + 50$$

$$= -17.50 + 50$$

$$= 32$$

and

$$\begin{aligned} T_2 &= 10\left(\frac{X_2 - E(X_2)}{\sigma_2}\right) + 50 \\ &= 10\left(\frac{45 - 51}{12}\right) + 50 \\ &= -5 + 50 \\ &= 45 \end{aligned}$$

Now that the T scores are available it is easy to determine which person is performing at a higher level. Since the T scores are on exactly the same scale, they may be added. Mr. Brown's total score is $T = 39 + 41 = 80$. Miss White's total score is $T = 32 + 45 = 77$. Even though their total raw scores are the same, it appears that Mr. Brown is doing slightly better than Miss White. If the standardized scores are normally distributed it is possible to determine the percentile rank of these scores. However, the determination of these percentile ranks will have to wait until the discussion on correlation is completed.

As these examples indicate, the normal distribution in conjunction with Theorems 7-1 to 7-7 gives rise to some important and powerful methods for the analysis of data. Of even greater importance is the combining of Theorems 6-1 and 6-4 with the normal probability distribution. The union of these theorems with the normal distribution gives rise to one of the most important theorems of mathematics and, incidentally, the *most important* theorem of statistical theory.

7-11 THE CENTRAL LIMIT THEOREM

According to Theorems 6-1 and 6-4, if $X_1, X_2, \ldots, X_N$ are identically and independently distributed with $E(X_i) = \mu_X$ and $\text{Var}(X_i) = \sigma_X^2$, then $T = X_1 + X_2 + \cdots + X_N$ is distributed with $E(T) = N\mu_X$ and $\text{Var}(T) = N\sigma_X^2$. To see what happens to T as N increases, consider three discrete probability distributions that have been studied in great detail in Chapter 6. In particular, consider

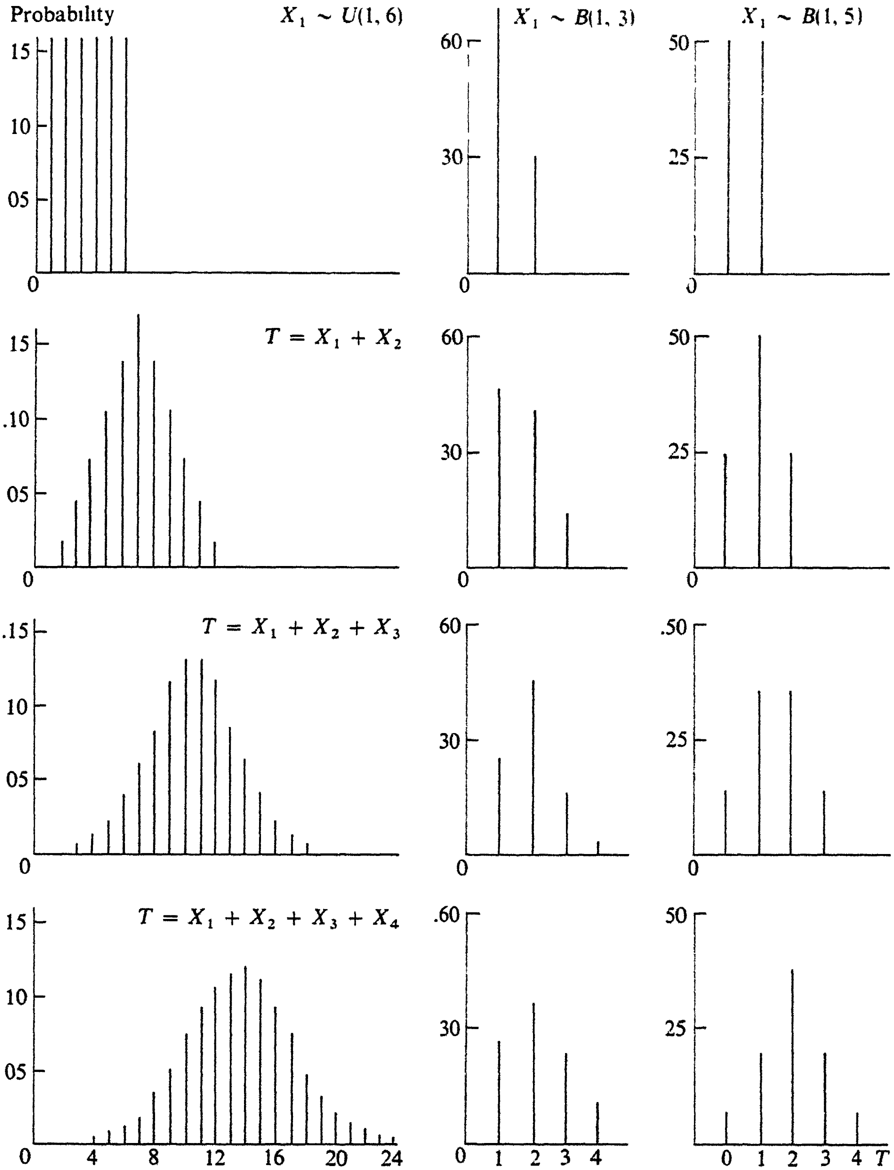
1. X, the uniform distribution $U(1, 6)$
2. Y, the Bernoulli distribution $B(1, .3)$
3. Z, the Bernoulli distribution with $B(1, .5)$

For these distributions,

1.	$E(X) = 3.5$	$\text{Var}(X) = 2.9167$	$\sigma_X = 1.71$
2.	$E(Y) = .30$	$\text{Var}(Y) = .21$	$\sigma_Y = .44$
3.	$E(Z) = .50$	$\text{Var}(Z) = .25$	$\sigma_Z = .50$

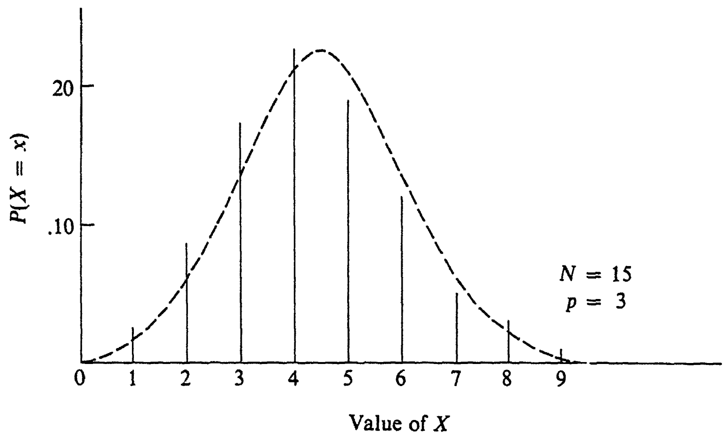
A graphic representation of these distributions is shown in the top row of Figure 7-14. Consider the variables $T_X = X_1 + X_2$, $T_Y = Y_1 + Y_2$, and $T_Z = Z_1 + Z_2$, which can be generated from the original variables. The distributions of these newly

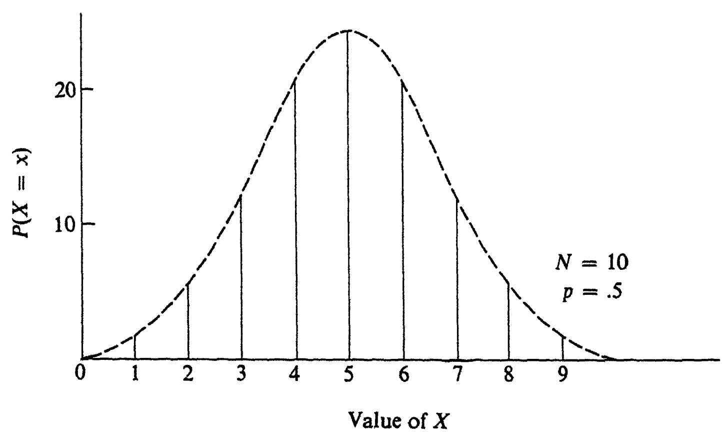
Figure 7-14. Illustration of the convergence of three discrete probability distributions to the normal form



constructed random variables are shown in the second row of Figure 7-14. In comparing the graphs of row 1 with the graphs of row 2, notice the change in forms of the distributions. Now consider the variables $T_X = X_1 + X_2 + X_3$, $T_Y = Y_1 + Y_2 + Y_3$, and $T_Z = Z_1 + Z_2 + Z_3$. The distributions of these random variables are shown in the third row of Figure 7-14. Again, particular attention should be paid to the change in form. Consider, now, the variables defined for $N = 4$; their

Figure 7-15. The similarity of the $B(15, .3)$ to the $N(4.5, 3\ 15)$ and the $B\ (10,\ 5)$ to the $N(5, 2.5)$.





corresponding graphs are shown in the figure. As before, consider the graphical presentation of the probability distributions associated with each of these new sets of variables.

The first point to be noted is that the range of T increases as N increases. Also, the expected values, the variances, and the standard deviations increase and correspond to the figures presented, *but more important is that the form of the distribution of T changes as N increases.* For the uniform variable X, T_X has a triangular distribution when $N = 2$. When $N = 3$, the distribution becomes flatter at the tails and near the expected value. When $N = 4$, the distribution might be confused with a normal distribution, though of course it cannot be normal since it represents the probability distribution of a discrete variable. As one might expect, the convergence upon normality becomes closer and closer as N becomes larger.

While the convergence of the two binomial variables to the normal is not as apparent as it is for the uniform distribution, this convergence also occurs. For the binomial, the convergence is much slower since it is dependent upon the value of N. This is illustrated in Figure 7-15 for $N = 15$ and $p = .30$ and $N = 10$ and $p = .50$. In both cases, the similarity of these two distributions to a normal distribution is striking.

It is the general convergence of *sums* of random variables to a probability distribution that is normal in form that makes the normal distribution such an important distribution for data analysis and research. The convergence illustrated in this section is not a function of these three distributions only. This convergence occurs for any random variable that is defined as the sum of other random variables. While the convergence has been illustrated for identically and independently distributed random variables, it could have been demonstrated for more complex variables that are not identically distributed and not necessarily independent. Since *most* of the applications encountered in the behavioral sciences make the assumption of identical and independent distributions, this is not necessarily a loss in generality. For that reason the central limit theorem is stated in the following restricted form.

Theorem 7-8. The central limit theorem

Let $X_1, X_2, \ldots, X_N$ have identically and statistically independent probability distributions with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Then the random variable $T = X_1 + X_2 + \cdots + X_N$ has an approximate normal probability distribution with $E(T) = N\mu$ and $\mathrm{Var}(T) = N\sigma^2$. As N approaches infinity, the distribution of T approaches the normal distribution.

In general, the normal approximation to most distributions is adequate for most behavioral research. Even in cases where it is not appropriate, it can serve as a useful first approximation. This will be illustrated for the three distributions of this section.

1. If $T = X_1 + X_2 + X_3 + X_4$ and X_i is $U(1,6)$, the exact probability that $T = 15$ is equal to .115. According to Theorems 6-1 and 6-4, $E(T) = 4\mu = 4(3.5) = 14$ and $\text{Var}(T) = 4\sigma^2 = 4(2.9167) = 11.67$. For the normal approximation,

$$\begin{aligned} P(T = 15) &= P(14.5 \leqslant T \leqslant 15.5) \\ &= P\left(\frac{14.5 - 14}{\sqrt{11.67}} \leqslant Z \leqslant \frac{15.5 - 14}{\sqrt{11.67}}\right) \\ &= P(.15 \leqslant Z \leqslant .44) \\ &= P(Z \leqslant .44) - P(Z \leqslant .15) \\ &= .6700 - .5596 \\ &= .1104 \end{aligned}$$

2. If $T = \sum_{i=1}^{15} X_i$ and X_i is Bernoulli with $p = .30$, the exact probability that $T = 8$ is equal to .0348. Since T is binomially distributed, $E(T) = Np = 15(.3) = 4.5$ and $\text{Var}(T) = Npq = 15(.3)(.7) = 3.15$. For the normal approximation,

$$\begin{aligned} P(T = 8) &= P(7.5 \leqslant T \leqslant 8.5) \\ &= P\left(\frac{7.5 - 4.5}{\sqrt{3.15}} \leqslant Z \leqslant \frac{8.5 - 4.5}{\sqrt{3.15}}\right) \\ &= P(1.69 \leqslant Z \leqslant 2.25) \\ &= P(Z \leqslant 2.25) - P(Z \leqslant 1.69) \\ &= .9878 - .9544 \\ &= .0334 \end{aligned}$$

3. If $T = \sum_{i=1}^{10} X_i$ and X_i is Bernoulli with $p = .50$, the exact probability that $T = 9$ is equal to .0098. For this distribution, $E(T) = 10(.5) = 5$ and $\text{Var}(T) = 10(.5)(.5) = 2.5$. For the normal approximation,

$$\begin{aligned} P(T = 9) &= P(8.5 \leqslant T \leqslant 9.5) \\ &= P\left(\frac{8.5 - 5}{\sqrt{2.5}} \leqslant Z \leqslant \frac{9.5 - 5}{\sqrt{2.5}}\right) \\ &= P(2.22 \leqslant Z \leqslant 2.85) \\ &= P(Z \leqslant 2.85) - P(Z \leqslant 2.22) \\ &= .9978 - .9868 \\ &= .0110 \end{aligned}$$

In every case, the approximation is quite good.

Finally, it should be noted that many biological variables, such as age of menarche (described in the opening paragraphs of this chapter), possess a normal distribution. Much of this tendency to normality is directly related to the central limit theorem and the properties of variables that are defined by $T = X_1 + X_2 + \cdots + X_N$.

7-12 THE NORMAL APPROXIMATION TO THE BINOMIAL

Since the binomial converges upon the normal, one should be able to use it to approximate binomial probabilities. To see that this approximation is adequate, suppose that an unfair coin with probability of heads equal to $\frac{1}{4}$ is tossed 200 times in succession. For this coin, the probability that X, the number of heads, is equal to 50 is, according to the binomial formula, given by

$$\begin{aligned} P(X = 50) &= \binom{200}{50} (p)^{50} (q)^{150} \\ &= \binom{200}{50} (\tfrac{1}{4})^{50} (\tfrac{3}{4})^{150} \end{aligned}$$

With considerable effort, this computation could be carried out. Fortunately, the computations are simplified if the normal distribution is used to *approximate* this probability. This approximation is valid for large N and if $Np > 5$ and $Nq > 5$. If these last two inequalities are not satisfied, the binomial distribution does not approach the normal distribution. When $Np < 5$ or $Nq < 5$ and N is large, the binomial approaches a different distribution, which is not discussed in this book. Since binomial variables are discrete, it is necessary to make appropriate scale adjustments. According to the continuity correction, $X = 50$ is equivalent to X falling within the interval 49.5 to 50.5.

The expected value of a binomial distribution is equal to Np. In this case, $Np = (200)(\frac{1}{4}) = 50$, which is greater than 5. Also, $Nq = (200)(\frac{3}{4}) = 150$ is greater than 5. Therefore, the binomial can be approximated by the normal. The variance is given by $Npq = (200)(\frac{1}{4})(\frac{3}{4}) = 37.5$. The standard deviation, which is the square root of the variance, is 6.12. Thus

$$\begin{aligned} P(X = 50) &= P(49.5 \leqslant X \leqslant 50.5) \\ &= P(X \leqslant 50.5) - P(X \leqslant 49.5) \\ &= P\left(Z \leqslant \frac{50.5 - 50}{6.12}\right) - P\left(Z \leqslant \frac{49.5 - 50}{6.12}\right) \\ &= P(Z \leqslant .08) - P(Z \leqslant -.08) \\ &= .532 - .468 \\ &= .064 \end{aligned}$$

This is a small probability of occurrence. Note that in only 6 times out of 100 would one expect to get the expected value of 50 heads. The occurrence of the expected value in the large number of trials has extremely low probability of happening.

7-13 PARAMETERS OF PROBABILITY DISTRIBUTIONS (CONTINUATION)

The expected value and the variance of a probability distribution are sometimes reported as though they were the parameters that characterize a distribution. For the normal distribution, these measures are indeed parameters, so that in the sequel a normal distribution with parameters μ and σ^2 will be denoted $N(\mu,\sigma^2)$. If a distribution is normal, then knowledge concerning these measures implies that a significant amount of information about the distribution is known. Probabilities can be determined and statements can be made about the spread and center of the distribution. Furthermore, all the information about its probability distribution is summarized in the two numbers μ and σ^2. As stated earlier, parameters are numbers used to characterize probability distributions. Except for simple gambling systems, these values are generally unknown. With respect to real-world studies and investigations, they must be estimated. The methods for estimating parameters are, for the most part, based upon simple random sampling procedures.

7-14 SUMMARY

In this chapter, probability distributions for continuous random variables were introduced. As has been emphasized throughout the discussion, behavioral data tend to be discrete. Since continuous probability models serve as excellent approximations to discrete probability models, they tend to dominate statistical theory.

Like discrete variables, continuous variables can be graphically represented to illustrate their main characteristics. If the probability distribution is defined in terms of mutually exclusive and exhaustive subsets, one can make a simple representation via a histogram or cumulative probability polygon. If the probabilities are specified by a mathematical equation, a smooth curve suffices. Like discrete variables, continuous variables can be symmetrical, positively skewed, negatively skewed, or bimodally distributed.

One special continuous distribution that serves as an excellent approximation for many behavioral variables is the normal distribution with parameters μ and σ^2, the expected value and variance of the distribution. This distribution is symmetrical and unimodal with an infinite range of possible values. For all practical purposes, 100 percent of the distribution is between $\pm 3\sigma$ of μ; 95 percent of the distribution is within $\pm 1.96\sigma$ of μ; and 68 percent of the distribution is within $\pm 1\sigma$ of μ.

A common practice of statistical theory is to transform variables to different scales that are more convenient for analysis. The four basic theorems associated with the expected value and variance of the resulting variable in terms of the expected value and variance of the original variable were summarized in Theorems 7-1 to 7-4. In essence, these important statistical theorems state that if X is a random variable with expected value μ_X and variance σ_X^2, then the variable Y produced by the

transforming equation $Y = A + BX$ has its expected value and variance given by $E(Y) = A + B\mu_X$ and $\mathrm{Var}(Y) = B^2\sigma_X^2$.

An important application of this transformation is related to the determination or probabilities of normally distributed variables. As was noted, if X is $N(\mu_X, \sigma_X^2)$, then $Z = (X - \mu_X)/\sigma_X$ is $N(0,1)$. As a result, one can determine the probability of any event for a variable that is normally distributed since the normal distribution with $E(Z) = 0$ and $\sigma_Z = 1$ has been tabled.

For example, suppose at a certain college it has been learned over many trials that sophomores can run the 100-yard dash in an average time of 12 seconds with a standard deviation of .6 seconds. If running times are normally distributed, then the proportion of men who can run the 100-yard dash in less than 11.2 seconds is given by

$$P(X \leqslant 11.2) = P\left(Z \leqslant \frac{11.2 - 12}{.6}\right) = P(Z \leqslant -1.33) = .092$$

If the variable is discrete, one need only make a correction for continuity before using the transformation equation. Thus, if the number of words an eighth grade student can read in a minute is approximately normally distributed with $E(X) = 225$ and $\sigma_X = 50$, then the proportion of students who read 300 or more words per minute is given by

$$\begin{aligned} P(X \geqslant 300) &= P\left(Z \geqslant \frac{299.5 - 225}{50}\right) \\ &= P\left(Z \geqslant \frac{74.5}{50}\right) = P(Z \geqslant 1.49) \\ &= 1 - P(Z \leqslant 1.49) = 1 - .9319 \\ &= .0681 \end{aligned}$$

Perhaps the most important reason for the special attention paid to the normal distribution is embodied in the central limit theorem. According to this theorem, if X_i, $i = 1, 2, \ldots, N$ are independently and identically distributed with $E(X_i) = \mu_X$ and $\mathrm{Var}(X_i) = \sigma_X^2$, then the new variable $T = X_1 + X_2 + \cdots + X_N$ tends to be normally distributed with $E(T) = N\mu_X$ and $\mathrm{Var}(T) = N\sigma_X^2$, as N increases. The remarkable thing about this theorem is that the X_i need not be normally distributed for the new variable T to approach a normal form. If a distribution X is symmetrical, the probability distribution of T approaches a normal form for relatively small N. However, some rather extreme distributions require a large N before T can be adequately described by a normal distribution. In general, when N exceeds 30, the similarity of T to a normal form is almost perfect.

One important special case of the central limit theorem concerns the convergence of binomially distributed variables to a normal form. If $Np > 5$ and $Nq > 5$, then X: {number of successes in N trials} can be approximated by the normal distribution

with $E(X) = Np$ and $\mathrm{Var}(X) = Npq$. Thus, if X: {number of families that have household pets} is binomially distributed with $p = .25$, then in a random sample of 300 households, the 95-percent range for X, corrected for continuity, is given by

$$(X_S - \tfrac{1}{2}) < X < (X_L + \tfrac{1}{2})$$
$$(Np - .5) - 1.96\sqrt{Npq} < X < (Np + .5) + 1.96\sqrt{Npq}$$
$$300(.25) - .5 - 1.96\sqrt{300(.25)(.75)} < X < 300(.25) + .5 + 1.96\sqrt{300(.25)(.75)}$$
$$74.5 - 14.7 < X < 75.5 + 14.7$$
$$60 \leqslant X \leqslant 90$$

Thus, in a household survey of 300 homes, one would expect to find pets in 60 to 90 of the homes surveyed.

Finally, it should be noted that variables that possess a hypergeometric distribution also tend to a normal form as the sample size increases. This means that hypergeometric probabilities can also be approximated by the normal distribution. For this approximation it is necessary that $np > 5$ and $nq > 5$. Furthermore, it is necessary that $E(X) = np$ and $\mathrm{Var}(X) = npq(N - n)/(N - 1)$, where N = universe size and $p = n(A)/N$, and where $n(A)$ = number of elements in the universe that have property A.

EXERCISES

***7-1.** For the data of Table 3-5 it was found that $P(M) = .672$. Suppose that 50 students are selected from each school and suppose it is found that the number of "more" responses is equal to the expected values, that is, $X_A = 50(.74) = 37$ and $X_B = 50(.52) = 26$. Using $p = \frac{2}{3}$ as a parameter, find

(*a*) $P(X_A \geqslant 37 | \text{School } A)$
(*b*) $P(X_B \geqslant 26 | \text{School } B)$
(*c*) What would these results suggest about the attitudes of the students at the two schools?
(*d*) What assumptions must you make in reaching this decision? Are they reasonable? Explain.

***7-2.** Consider the discrete random variable of Table 6-3. If 10 subjects are to be tested and if $X_1, X_2, \ldots, X_{10}$ represent the number of cigarettes they can correctly identify, provided that correct matching of cigarette with brand name is a chance event, find, for $T = X_1 + X_2 + \cdots + X_{10}$

(*a*) $P(T \geqslant 10)$
(*b*) $P(T \geqslant 13)$
(*c*) $P(T \geqslant 16)$
(*d*) Suppose $T = 15$; would you conclude that they really knew the brands of cigarettes? Explain.

***7-3.** What is the central 68 percent range of T for Exercise 6-4? What assumptions have you made in determining this range? Inspection of the probability distribution of X (Exercise 6-1) shows that the modal attendance is 5 movies per month. Find the probability for $T = X_1 + X_2 + X_3 + X_4 + X_5$, that $T \geqslant 25$.

***7-4.** If, for college male sophomores, the running time in seconds for the 100-yard dash is $N(12, .8^2)$, find

(*a*) $P(X \geqslant 14)$
(*b*) The 75th percentile of the distribution
(*c*) The 95 percent central range in running times

***7-5.** If two men are selected to run the 100-yard dash, and if $T = X_1 + X_2$ = their total running time, find

(*a*) $P(T \geqslant 28)$
(*b*) The 75th percentile of the distribution of T
(*c*) The 95 percent central range in total running times

***7-6.** In a survey of 200 families in Los Angeles, it was found that the amount of money spent on food for a family of four averages \$37 per week with a standard deviation of \$6.

(*a*) What is the probability that in a four-week period, a family of four will spend more than \$160 on food?
(*b*) Find Q_1 and Q_3 for a four-week spending period.
(*c*) Since economic variables tend to be positively skewed, what can you say about the median amount of money spent on food for one week?
(*d*) What can you say about the median amount of money spent on food for four weeks? What is the rationale for your decision?

7-7. Miss Green and Miss Black, when applying for a secretarial job, took a typewriting and shorthand test. Their scores and the norms for the test in words per minute are as follows:

Test	*Typing*	*Shorthand*
Expected score	50	80
S.D.	4	15
Miss Green	52	90
Miss Black	56	86

Which of the two girls would you hire? Why?

***7-8.** If male heights in inches are $N(68, 4^2)$, find

(*a*) $E(5 + 6X)$
(*b*) $\mathrm{Var}(5 + 6X)$

***7-9.** If, for the height distribution of Exercise 7-8, three men are selected at random and if their heights are denoted by X_1, X_2, and X_3, find

(a) $E(7 + 2X_1 + 3X_2 - 4X_3)$
(b) $\mathrm{Var}(7 + 2X_1 + 3X_2 - 4X_3)$
(c) $P(7 + 2X_1 + 3X_2 - 4X_3 > 80)$

7-10. Forty percent of incoming freshmen at a large west coast university drop out of school before the completion of their first year.

(a) If a sample of 100 students is followed in their academic career as part of a longitudinal study, find the probability that more than 70 return for the beginning of their sophomore year. What does this probability suggest about the assumption that 40 percent drop out?
(b) What is the 95-percent range for the number of returnees for the beginning of the sophomore year?
(c) What assumptions have you made? Are they reasonable? Explain.

8 INTRODUCTION TO THE STATISTICAL THEORY OF ESTIMATION

A poll of 672 male senior and graduate students at San Francisco State College showed yesterday that 67 percent would refuse a draft call.

The poll was conducted on campus last week by a small newly formed students group called the Anti-Draft Union. The 672 students responding in the survey represent about 13 percent of the 4920 male senior and graduate student enrollment.

Survey questionnaires were passed out on an informal basis and returned to tables manned by the Anti-Draft Union in front of the college commons and the library. There were no formal controls. . . .

By permission, from *San Francisco Chronicle,* San Francisco, California, March 26, 1968.

8-1 RANDOM SAMPLING

As was suggested in previous chapters, the populations that are of greatest interest to the behavioral scientist are hypothetical or abstract in nature. For these populations, the behavioral scientist would not be likely to have all the information that exists about a characteristic possessed by the elements of the population. As a consequence, sampling becomes a necessity. Since samples are selected to illuminate and simplify decision making about populations, one wants to be sure that the information collected gives a good representation of the population. One of the best ways for ensuring a good population representation is to select a random sample. One of the conditions that must be satisfied in selecting a random sample is that each of the elements in the population should have an equal probability of being selected. Furthermore, the sampling should ensure independence of selection. These two conditions essentially define a sampling procedure that is called simple random sampling. According to definition, a sampling procedure is said to represent *simple random sampling* if

1. Every element of the sample is selected independently of all other elements of the population.
2. Probability of selection into the sample is equal for all elements in the population.

According to this definition, the survey of attitudes toward opposition to the draft at San Francisco State College cannot possibly be based on a random sample of the college students of direct interest. As was reported, "Survey questionnaires were passed out on an informal basis . . . ." Random sampling, on the other hand, is exceptionally formal. For random sampling, *all* students must have an equal probability of being selected. In the San Francisco State College survey, only students with strong commitments who were fortunate enough to be informally contacted had any sort of chance of responding. For the noncontacted individual, the probability of responding was essentially 0. Furthermore, the independence of selection is suspect since friends of similar beliefs most likely influenced one another in responding to the questionnaire.

As another example of an attitude survey consider a political scientist who wanted to make a study of the public's reaction to the last primary election in California and suppose that he wished to find out if the people were in agreement with the declared policy statements of the Republicans or Democrats. To obtain a sample to interview, he could decide to select people as they walked past the fountain in the student plaza of the University of California. They could then be asked to state their opinions on the stated policies of the two political parties. It takes little imagination to realize that this sampling procedure would not produce a random sample of California citizens because the probability of being included in the sample is dependent upon walking by the area where the poll-taker has been stationed to ask his questions. Most people sampled will be students or employees of the university. Their views will not be representative of those of the vast majority of California citizens.

Suppose a study is to be conducted in a large school district in which a classroom of third-grade children is to be selected In all probability, this method of sampling would not generate a random sample of all third-grade children of the school district since schools frequently use ability groupings; furthermore, as is well known, schools in different neighborhoods are significantly different from one another in socioeconomic factors and in the effectiveness and commitment of the teachers. Unavoidably, children in different schools reflect these neighborhood differences.

Since most studies involve the collection of many bits of information on various topics from each element in the sample, it is frequently advisable to transfer the data either to cards or to tapes. For most small studies, computers will not be used and simple 3×5 cards are sufficient for recording data. However, with samples greater than 200, IBM cards are generally most appropriate. When a researcher must make a choice from 3×5 cards, IBM cards, or tapes, it is not unwise for him to consult with a statistician before proceeding with his data analysis.

Finally, it must be emphasized that currently employed statistical methods are dependent upon random sampling and the assumptions underlying this type of sampling. If the assumptions of this sampling procedure are not satisfied, the statistical methods to be presented in this book should not be used. With random samples one can obtain precise estimates of population parameters. Furthermore, one can make inferences about the populations from which random samples are selected. One of the best ways for obtaining a random sample is to use random numbers listed in a random number table. This procedure is described in Section 8-2.

8-2 SELECTING A SAMPLE FROM A FINITE POPULATION

In this section, simple random sampling from a finite population is shown. Consider a universe of 500 students who were given a test of reading ability. The scores on the test are listed in Table 8-1. In place of names, each student has been identified by a number from 001 to 500. For illustrative purposes, assume that the test had never been given; a sample of 25 students can be selected whose scores will be used to estimate the expected score and standard deviation of the test. The simplest way to obtain a random sample for this population is to open a book of random numbers to any page at random, and then with the eyes closed to place an index finger on the page at random, and then to record three-digit numbers as they appear on the page, excluding all numbers greater than 500. This process should be continued until 25 three-digit random numbers have been selected. Table A-5 of Appendix A, which is a shortened version of a more complete table of random numbers, can also be used in exactly this way. The following set of numbers were selected in exactly this manner from the table of random numbers in the appendix: R: {467, 737, 572, 931, 407, 141, 623, 137, 267, 967, 053, 475, 250, 603, 385, 354, 949, 150, 826, 903, 757, 473, 416, 461, 339, 617, 305, 000, 810, 909, 147, 190, 432, 624, 087, 731, 494, 907, 300, 664, 666, 075, 110, 198, 193, 493, 388, 499, 751, 803}.

Table 8-1. Scores obtained by 500 students on a reading test (mean and variance unknown).

Student number	*Score*	*Student number*	*Score*	*Student number*	*Score*	*Student number*	*Score*	*Student number*	*Score*
001	55	039	37	076	52	113	41	151	47
002	51	040	62	077	53	114	50	152	48
003	65			078	56	115	41	153	57
004	60	041	56	079	47	116	62	154	85
005	64	042	54	080	39	117	36	155	53
006	59	043	36			118	40	156	56
007	62	044	48	081	51	119	57	157	58
008	35	045	47	082	34	120	57	158	49
009	43	046	54	083	47			159	47
010	64	047	56	084	45	121	34	160	58
		048	59	085	80	122	61		
011	45	049	52	086	63	123	62	161	38
012	36	050	42	087	49	124	31	162	46
013	40			088	41	125	54	163	66
014	50	051	51	089	52	126	39	164	43
015	64	052	25	090	48	127	47	165	62
016	32	053	46			128	56	166	60
017	49	054	45	091	45	129	51	167	54
018	37	055	44	092	31	130	59	168	54
019	60	056	45	093	55			169	64
020	53	057	40	094	48	131	35	170	41
		058	45	095	55	132	42		
021	32	059	58	096	46	133	52	171	55
022	38	060	52	097	56	134	63	172	59
023	56			098	41	135	44	173	38
024	46	061	67	099	44	136	44	174	59
025	36	062	48	100	54	137	48	175	44
026	52	063	56			138	41	176	48
027	42	064	41	101	74	139	38	177	51
028	73	065	38	102	45	140	47	178	43
029	50	066	47	103	44			179	58
030	39	067	46	104	63	141	58	180	51
		068	68	105	50	142	52		
031	58	069	55	106	45	143	33	181	50
032	54	070	30	107	50	144	46	182	52
033	45			108	48	145	54	183	43
034	53	071	30	109	34	146	56	184	38
035	60	072	62	110	54	147	39	185	45
036	37	073	61			148	49	186	49
037	47	074	54	111	50	149	50	187	56
038	52	075	48	112	64	150	58	188	45

Table 8-1. Scores obtained by 500 students on a reading test (mean and variance unknown) *(Continued).*

Student number	*Score*	*Student number*	*Score*	*Student number*	*Score*	*Student number*	*Score*	*Student number*	*Score*
189	54	225	44	261	58	298	64	334	45
190	68	226	47	262	48	299	46	335	43
191	51	227	32	263	54	300	47	336	79
192	45	228	52	264	49	301	47	337	38
193	58	229	44	265	48	302	52	338	52
194	42	230	54	266	34	303	58	339	59
195	35	231	39	267	49	304	52	340	62
196	60	232	44	268	39	305	48	341	67
197	36	233	60	269	37	306	66	342	49
198	45	234	41	270	36	307	61	343	49
199	45	235	39	271	48	308	54	344	68
200	58	236	45	272	48	309	46	345	49
201	49	237	45	273	54	310	60	346	55
202	55	238	64	274	56	311	45	347	64
203	59	239	40	275	52	312	52	348	50
204	56	240	64	276	62	313	48	349	45
205	79	241	55	277	46	314	61	350	39
206	59	242	57	278	63	315	62	351	63
207	60	243	58	279	55	316	49	352	38
208	60	244	53	280	33	317	33	353	40
209	45	245	64	281	57	318	40	354	62
210	52	246	49	282	53	319	66	355	54
211	45	247	34	283	50	320	60	356	52
212	47	248	55	284	51	321	50	357	44
213	51	249	41	285	59	322	52	358	57
214	61	250	54	286	44	323	53	359	48
215	48	251	43	287	38	324	59	360	35
216	45	252	34	288	45	325	47	361	31
217	40	253	64	289	58	326	61	362	58
218	60	254	32	290	56	327	42	363	53
219	60	255	70	291	46	328	51	364	55
220	39	256	41	292	49	329	55	365	24
221	52	257	57	293	43	330	60	366	46
222	63	258	52	294	60	331	51	367	22
223	41	259	30	295	70	332	49	368	54
224	48	260	43	296	51	333	54	369	40
				297	34			370	51

Table 8.1 Scores obtained by 500 students on a reading test (mean and variance unknown) *(Continued)*

Student number	*Score*	*Student number*	*Score*	*Student number*	*Score*	*Student number*	*Score*	*Student number*	*Score*
371	51	398	37	424	47	451	48	478	44
372	46	399	63	425	48	452	64	479	40
373	41	400	47	426	56	453	40	480	37
374	46			427	57	454	55		
375	55	401	67	428	30	455	31	481	60
376	50	402	56	429	44	456	47	482	40
377	37	403	52	430	43	457	50	483	56
378	32	404	50			458	54	484	54
379	40	405	42	431	55	459	43	485	51
380	36	406	45	432	44	460	46	486	55
		407	54	433	44			487	55
381	41	408	31	434	49	461	45	488	53
382	48	409	48	435	52	462	51	489	61
383	42	410	46	436	51	463	48	490	61
384	42			437	20	464	41		
385	37	411	58	438	54	465	40	491	58
386	55	412	47	439	49	466	54	492	52
387	47	413	32	440	55	467	50	493	47
388	46	414	60			468	56	494	61
389	60	415	40	441	50	469	26	495	52
390	60	416	33	442	54	470	70	496	49
		417	44	443	59			497	38
391	57	418	44	444	54	471	47	498	61
392	62	419	39	445	42	472	41	499	41
393	45	420	33	446	42	473	40	500	62
394	60			447	50	474	55		
395	79	421	55	448	61	475	77		
396	52	422	44	449	57	476	50		
397	33	423	55	450	47	477	47		

Eliminating all random numbers that are not associated with a student, the behavioral scientist selects the following 25 numbered students for inclusion in the sample: S: {467, 407, 141, 137, 267, 053, 475, 250, 385, 354, 150, 473, 416, 461, 339, 305, 147, 190, 432, 087, 494, 300, 075, 110, 198}. Note that in this listing no student appears more than once. If a student had been selected twice, he would have been replaced by another student. This produces a sampling scheme that involves sampling without replacement. The scores associated with these random numbers are as follows: {50, 54, 58, 48, 49, 46, 77, 54, 37, 62, 58, 40, 33, 45, 59, 48, 39, 68, 44, 49, 61, 47, 48, 54, 45}.

As the numbers stand, they impart little information about the expected value, the standard deviation, or the form of the distribution. Intuition would suggest that the sample should mimic the population and should be used in a manner analogous to a population.

Just as when studying probability distributions, it is convenient to convert sample values into a small set of numbers, especially if the sample size is large. As with continuous variables, one might like to associate probabilities with certain mutually exclusive subsets of the sample, and then possibly to draw a diagram such as a histogram or a line graph to represent the sample and to use it as a partial description of the population.

8.3 FREQUENCY TABLES

As a first step to data reduction, it is informative to determine the sample range. By definition, the *sample range* is the difference between the largest observed value X_L and the smallest observed value X_S. Thus, for the data of Section 8-2,

$$\hat{R} = X_L - X_S = 77 - 33 = 44$$

where $\hat{R}$ equals the sample range and is an estimate of R, the population range. Numbers such as $\hat{R}$, computed from a sample and used as an estimator of a population value, are called statistics. Thus, any measure made on a sample represents a *statistic*.

Since the scores are discrete, one could list the scores and count the numbers of scores at each value. If each frequency of occurrence is divided by the total sample size, the resultant ratios are the observed relative frequencies of occurrence. One could then use these relative frequencies as estimates of the probabilities of occurrence. From the relative-frequency interpretation of probability, it is known that if the sample is allowed to increase in size, the relative frequencies will converge to the true population probabilities. With these estimates, one could estimate the expected value and the standard deviation by using the relative frequencies as substitutes for the probabilities. Furthermore, one could prepare an appropriate graph to serve as an approximation to the population graph. In general, most researchers do not follow this approach, not because it is incorrect but because it is time-consuming. Instead, short-cut methods are used that give about the same information in a more meaningful fashion.

For the purposes of data reduction, it is convenient to group the observed outcomes into 8 to 12 equal-width intervals. Fewer than 8 intervals is to be avoided, since this tends to distort and obscure relationships. More than 12 is to be avoided because it makes mathematical calculation unnecessarily complicated. For the observed sample, a convenient number of intervals is 10, each containing 5 discrete score values. These are shown in Table 8-2 in sets of 5 discrete values. The second column of this table lists what are called the *apparent scores*. These scores aid in tallying or counting the scores. The tallies are also shown in Table 8-2, along with

Table 8-2. Table of tallies for the observed sample of 25 elements selected at random from Table 8-1.

Raw scores	*Apparent intervals*	*Tallies*	*Frequency*	*Relative frequency*
30, 31, 32, 33, 34	30–34	/	1	04
35, 36, 37, 38, 39	35–39	//	2	.08
40, 41, 42, 43, 44	40–44	//	2	.08
45, 46, 47, 48, 49	45–49	~~////~~ ////	9	.36
50, 51, 52, 53, 54	50–54	////	4	.16
55, 56, 57, 58, 59	55–59	///	3	.12
60, 61, 62, 63, 64	60–64	//	2	.08
65, 66, 67, 68, 69	65–69	/	1	.04
70, 71, 72, 73, 74	70–74		0	.00
75, 76, 77, 78, 79	75–79	/	1	.04
			25	1.00

the frequencies and relative frequencies. Note that the relative frequencies are low at the extremes and then increase to give a modal class of occurrence at 45 to 49. By convention, the modal class is defined as the class or interval with the greatest relative frequency. The *mode* of the sample is the value with the greatest relative frequency. If the data have been tabulated, then the mode is defined as the midpoint of the interval with the greatest relative frequency. In this example, the mode equals 47. This is an estimate of the mode of the 500 scores.

Combining the scores into mutually exclusive and exhaustive intervals has, in essence, created a hypothetical "continuous" variable. To make the conversion "complete," assume the continuity correction that states that a score X should be treated as representing a score in the interval $X \pm \frac{1}{2}U$, where $\frac{1}{2}U$ is half the unit of measurement. In this case, $U = 1$. Thus, a score of 45 is considered as representing an observation on a continuous variable with a value in the range $45 \pm \frac{1}{2}$. These intervals are called *true intervals*. It should be realized that these are not true intervals in the usual meaning of that term, but intervals that help to make arithmetic easier and that give results closer to those that would be obtained without grouping.

From the table of tallies one can now set up a table that is popularly referred to as a frequency or *relative-frequency table*. It should be noted that the end points of the true intervals can never occur in practice. For example, none of the students could have earned a score of 29.5, 34.5, 54.5, ..., or 79.5.

If one is measuring heights, and the first interval is defined by $60\text{–}63\frac{7}{8}$ inches and the ruler has notches at every $\frac{1}{8}$ inch, then the true interval extends from $59\frac{15}{16}$ to $63\frac{15}{16}$. Notice that a person's height would never be reported as being $59\frac{15}{16}$ or $63\frac{15}{16}$ inches since the ruler is not marked at such values. Instead, all people counted

in that interval will have heights of $\{60, 60\frac{1}{8}, 60\frac{2}{8}, 60\frac{3}{8}, \ldots, 63\frac{6}{8}, 63\frac{7}{8}\}$ inches A person with height 64 inches would be counted in the next higher interval.

Frequently, variables are reported to the last unit of measurement. An example is age. Most people give their age at their last birthday. For such a variable, the following kind of grouping is used: 15–19, 20–24, 25–29, etc. For this kind of variable the true limits are 15.00000 ... to 19.99999 ..., and so forth.

These examples indicate that considerable care must be taken in determining the true limits for any grouping that one uses to simplify arithmetic and induce data reduction for a statistical study.

8-4 SAMPLE HISTOGRAMS, RELATIVE-FREQUENCY POLYGONS, AND CUMULATIVE RELATIVE-FREQUENCY POLYGONS

A relative-frequency histogram or cumulative relative-frequency polygon can be drawn in much the same manner as that used for probability distributions. The relative frequencies can be treated as probabilities. They can be plotted in terms of the true intervals or in terms of the midpoints of the intervals. The following set of graphs is drawn using the figures of Table 8-3. For the cumulative relative-frequency polygon, points are plotted at the upper true limits of the intervals.

Table 8-3. Frequency table for the observed sample of 25 elements selected at random from Table 8-1.

Apparent intervals	*True intervals*	*Midpoint of the interval*	*Frequency*	*Relative frequency*	*Cumulative relative frequency*
30–34	29.5–34.5	32	1	.04	.04
35–39	34.5–39.5	37	2	.08	.12
40–44	39.5–44.5	42	2	.08	.20
45–49	44.5–49.5	47	9	.36	56
50–54	49.5–54.5	52	4	.16	.72
55–59	54.5–59.5	57	3	.12	.84
60–64	59.5–64.5	62	2	.08	.92
65–69	64.5–69.5	67	1	.04	.96
70–74	69.5–74.5	72	0	.00	.96
75–79	74.5–79.5	77	1	.04	1.00
			25	1.00	

8-5 THE MEDIAN AND OTHER PERCENTILES OF A SAMPLE

Quartiles and percentiles of a sample may be determined in much the same way as they were for probability distributions. These definitions parallel those of a population. For example, the median of a sample is the sample value that partitions the

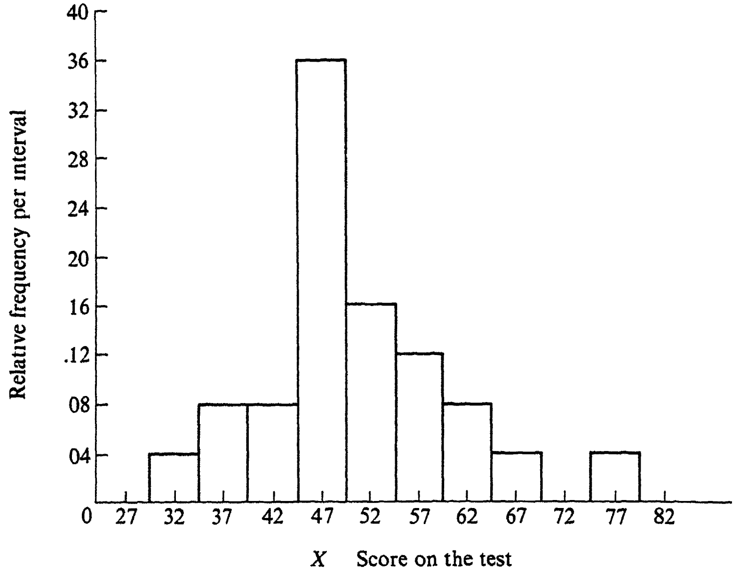
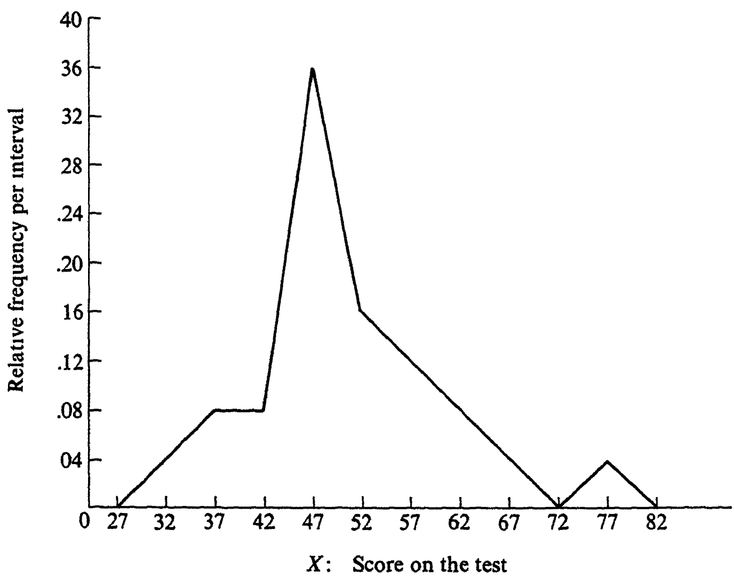


Figure 8-1. Histogram of the 25 test scores selected at random from Table 8-1

Figure 8-2. Relative-frequency polygon of the 25 test scores selected at random from Table 8-1.



sample into two parts such that half is below the median value and half is above the median value. Percentiles and quartiles are defined in analogous manner.

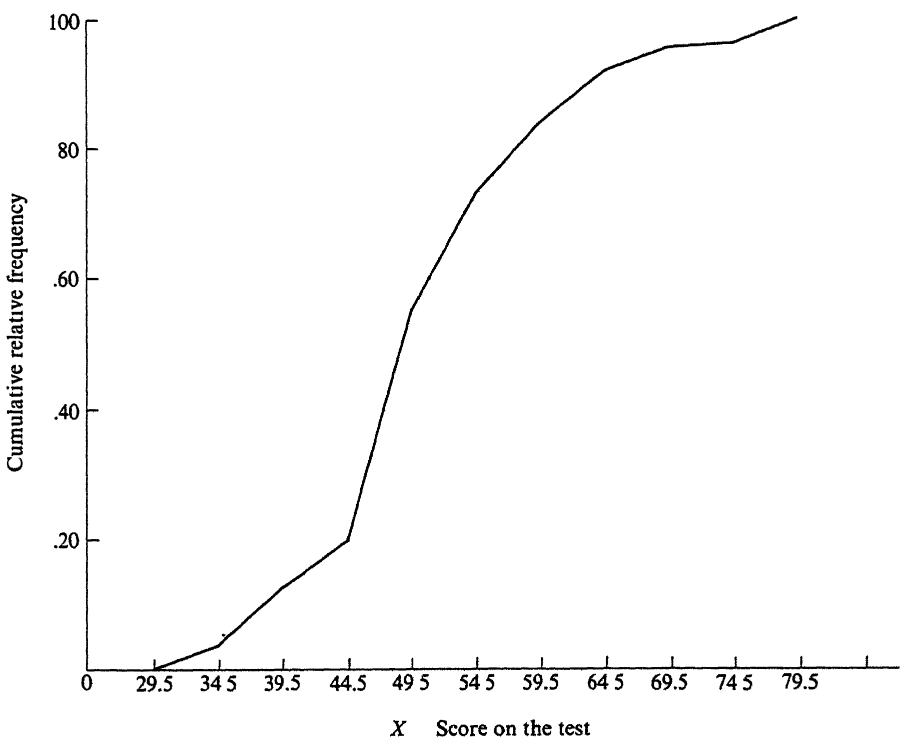
If the data have not been grouped and if the number of observations is odd, the median of the sample is taken as the middle ordered value. If the number of observations is even, the median of the sample is set equal to half the sum of the two middle ordered values. In the following two examples, the medians are both equal to 23.6:

Sample 1: {3.7, 4.9, 13.4, 23.6, 31.7, 32.4, 38.2}
Sample 2: {1.5, 12.3, 17.6, 29.6, 31.8, 40.7}

For grouped data, the median and other percentiles are simple to compute using linear interpolation. In order to determine the median for the 25 grouped scores, set up a table like Table 8-4. From this table a simple ratio equation of the differences can be created. For Table 8-4, this ratio equation is given by

$$\frac{(44.5 + \Delta) - 44.5}{49.5 - 44.5} = \frac{.50 - .20}{.56 - .20}$$

Figure 8-3. Cumulative relative-frequency polygon of the 25 test scores selected at random from Table 8-1.



or

$$\frac{\Delta}{5} = \frac{.30}{.36}$$

which gives

$$\Delta = \frac{5(.30)}{.36} = \frac{1.50}{.36} = 4.1666$$

Thus, the sample median $\hat{M}$ is given by

$$\hat{M} = 44.5 + 4.1666 = 48.6666 = 48\tfrac{2}{3}$$

Table 8-4. Linear interpolation table for estimating the median of the 25 test scores selected at random from Table 8-1.

		Value of X	Percent of scores below X		
5	Δ	44.5	.20	30	.36
		$44.5 + \Delta$	.50		
		49.5	.56		

For the ungrouped data, the median equals 49. Clearly, the grouped median value is quite close to the true sample median. Note that the median of the population is still unknown. To determine it, all 500 cases would have to be tallied. However, a usable estimate of the unknown population median is the sample median of $48\frac{2}{3}$. For all practical purposes, one would predict that 250 test scores are below $48\frac{2}{3}$ and 250 are above it.

Table 8-5. Linear interpolation table for estimating the 80th percentile of the 25 test scores selected at random from Table 8-1.

		Value of X	Percent of scores below X		
5	Δ	54.5	.72	.08	.12
		$54.5 + \Delta$	.80		
		59.5	.84		

In a like manner, the 80th percentile can be computed by setting up a table such as Table 8-5. For this table the simple ratio equation is given by

$$\frac{\Delta}{5} = \frac{.08}{.12}$$

from which it follows that

$$\Delta = \frac{5(.08)}{.12} = \frac{.40}{.12} = 3.3333$$

Thus, the 80th percentile, $\hat{P}_{80}$, is given by

$$\hat{P}_{80} = 54.5 + 3.3333 = 57.83333 = 57\tfrac{5}{6}$$

Again, it should be noted that the population 80th percentile is still unknown and may never be known. However, a good estimate of it is $57\frac{5}{6}$. For all practical purposes, 80 percent of the test scores are below this value. As to how many test scores are in reality below $57\frac{5}{6}$, this will remain a mystery forever. Of course, these percentiles and others could be read with ease from the cumulative relative-frequency polygon in exactly the same manner illustrated for the cumulative probability polygon.

8-6 THE MEAN OR AVERAGE VALUE OF A SAMPLE

Restricting the discussion to discrete variables, we find that the expected value of a population or probability distribution is given by

$$E(X) = \sum_{k=1}^{K} x_k P(X = x_k)$$

If one looks upon the sample as a miniature population, then the sample has an *expected value*. If the sample is denoted by X: $\{x_1,x_2,x_3, \ldots, x_i, \ldots, x_N\}$ and if all scores in the sample are given equal weight or are treated as being equally likely, the probability of any observation is, by Theorem 3-1, given by

$$P(X = x_i) = \frac{1}{N}$$

Thus, the *sample expected value* is given by

$$\bar{X} = \sum_{i=1}^{N} x_i \frac{1}{N}$$

which is simply the sample average of the sample values and is generally denoted by $\bar{X}$. Since this is such an important statistical measure, it is repeated as a definition. The *sample average* or *sample mean* is the sum of all the observations divided by the number of observations.

Thus, the average value of the following sample of five items, X: {3,8,14,6,9}, is equal to

$$\bar{X} = \frac{3+8+14+6+9}{5} = \frac{40}{5} = 8$$

If one "imagines" the sample to be a population and if each score has an equal probability of $\frac{1}{5}$, then the sample average could be computed in the same way as the expected value. This is illustrated in Table 8-6.

Table 8-6. Computation of a sample mean as the expected value of a uniform distribution with outcomes as shown.

Value of X	$P(X)$	$x_i P(X=x_i)$
3	$\frac{1}{5}$	$\frac{3}{5}$
6	$\frac{1}{5}$	$\frac{6}{5}$
8	$\frac{1}{5}$	$\frac{8}{5}$
9	$\frac{1}{5}$	$\frac{9}{5}$
14	$\frac{1}{5}$	$\frac{14}{5}$
Total	1	$\frac{40}{5} = 8 = \bar{X}$

As this discussion suggests, the sample average is related to a sample in exactly the same way as an expected value is to a population or probability distribution.

If a sample is large, the determination of the sample average might become a formidable task. If the data have been tallied and grouped as in a frequency table, the determination of the sample average is significantly easier. This simplification is achieved by assuming that every value of X is equal to the midpoint of the interval in which it is tallied. While it may appear that such an assumption results in a loss of information, the loss is overcompensated for by the simplification of the arithmetic. In addition, the discrepancy in the resulting computations is generally very small and of no practical significance. Furthermore, it should be emphasized that a sample is only one of many samples that could be drawn from a population. This means that a sample average is not an absolute measure. While it may be absolute for its particular sample, it is not for the population. Another sample drawn either at the same time or at a different time is almost certain to have a different average value. Because of this, one should get in the habit of referring to a sample average as a reasonable estimate of the population expected value. Thus, for example, if a sample average is 36.8, one should not hesitate to say that the expected value of the population is close to 37 in numerical value.

Table 8-7. Illustration of the computation of the sample mean for the grouped data of Table 8-3.

Midpoints of original intervals x_k	*Frequency of the interval* f_k	$x_k f_k$	*Code scale* y_k	$y_k f_k$	$y_k^2 f_k$
32	1	32	−3	−3	9
37	2	74	−2	−4	8
42	2	84	−1	−2	2
47	9	423	0	0	0
52	4	208	1	4	4
57	3	171	2	6	12
62	2	124	3	6	18
67	1	67	4	4	16
72	0	0	5	0	0
77	1	77	6	6	36
	25	1260		17	105

Table 8-7 is an illustration of the computations required to determine the sample average and the standard deviation (defined in Section 8-8) when data have been grouped. Let $x_1, x_2, \ldots, x_k, \ldots, x_K$ represent the midpoints of the K intervals and assume that all values of X in any one interval are equal to the midpoint value; the sample average is then given by

$$\bar{X} = \frac{32 + (37 + 37) + (42 + 42) + \cdots + (62 + 62) + 67 + 77}{25}$$

$$= \frac{32(1) + 37(2) + 42(2) + \cdots + 62(2) + 67(1) + 72(0) + 77(1)}{25}$$

$$= \tfrac{1260}{25} = 50.4$$

Thus, the population expected value is close to 50 in numerical value.

The arithmetic for this example is summarized in column 3 of Table 8-7. As this example suggests, the sample average for grouped data is given by

$$\bar{X} = \sum_{k=1}^{K} \frac{f_k x_k}{N} = \frac{1}{N} \sum_{k=1}^{K} f_k x_k$$

where x_k is the midpoint of the kth interval and f_k is the frequency in the kth interval.

As can be seen, the sample average can also be derived by multiplying the midpoints of the intervals by the relative frequencies and adding. If the relative frequencies are denoted by $\hat{p}_k = f_k/N$, then

$$\bar{X} = \sum_{k=1}^{K} \frac{f_k}{N} x_k = \sum_{k=1}^{K} \hat{p}_k x_k$$

This last form for computation of the sample average is analogous to the definition of the expected value of a discrete random variable except that the probabilities are replaced by relative frequencies.

8-7 COMPUTING FORMULA FOR THE AVERAGE ON A TRANSFORMED SCALE

In the discussion on probability distributions, it was shown that by adding a constant to all values a variable can assume or by multiplying every value by a constant, the expected value was affected in exactly the same way. For example, if every value of a random variable was decreased by 47, the expected value was also decreased by 47, and if every value was multiplied by $\frac{1}{5}$, the expected value was also multiplied by $\frac{1}{5}$. Since the sample average bears the same relationship to a sample as an expected value does to a population, the sample average should also be affected in exactly the same way, and as might be expected, it is.

A very useful transformation for computing the sample average is summarized in the following equation:

$$Y = \frac{X - \alpha}{\beta}$$

where α is the midpoint of any interval. Experience has shown that if the sample distribution is unimodal with its modal class near the center of the sample distribution, a wise choice for α is the midpoint of the modal class. If β is chosen equal to the width of any interval, then the arithmetic becomes very easy.

For the example, $\alpha = 47$ and $\beta = 5$, so that the transforming equation is

$$Y = \frac{X - 47}{5}$$

When $X = 32$,

$$Y = \frac{32 - 47}{5} = \frac{-15}{5} = -3$$

when $X = 37$,

$$Y = \frac{37 - 47}{5} = \frac{-10}{5} = -2$$

when $X = 42$, $Y = -1$; when $X = 47$, $Y = 0$; etc.

This new scale is shown as column 4 of Table 8-7. The average on this scale is given by

$$\bar{Y}=\sum_{k=1}^{K}\frac{f_k y_k}{N}=\frac{17}{25}=.68$$

To determine $\bar{X}$ one need only solve the transforming equation for X. Thus

$$5Y=X-47$$
$$X=5Y+47$$

When $Y=\bar{Y}=.68$, $X=\bar{X}$ is given by

$$\bar{X}=5(.68)+47=3.4+47=50.4$$

Thus, the sample average is 50.4 and the population expected value is estimated to be about 50.

In summary, to determine the sample average for grouped data one first transforms the midpoint values by the following equation:

$$Y=\frac{X-\alpha}{\beta}$$

where α = midpoint of any interval and β = width of all intervals. Then one computes

$$\bar{Y}=\sum_{k=1}^{K}\frac{f_k y_k}{N}$$

and then $\bar{X}$ is found by the following decoding equation:

$$\bar{X}=\beta\bar{Y}+\alpha$$

This last equation has an expected value theorem analog based on Theorems 7-1 and 7-3.

Theorem 8-1

If $E(Y)=\mu_y$ and if $X=\beta Y+\alpha$, then the $E(X)$ is given by

$$E(X)=\beta E(Y)+\alpha$$

Proof. By definition, $E(Y)=E(\beta Y+\alpha)=E(\beta Y)+E(\alpha)$. Since β and α are constants, it follows from Theorems 7-1 and 7-3 that

$$E(X)=\beta E(Y)+\alpha$$

This completes the proof.

8-8 THE VARIANCE AND STANDARD DEVIATION OF A SAMPLE

Proceeding as was done for the sample average, one could define a sample variance as follows:

$$\hat{S}^2 = \sum_{i=1}^{N} (x_i - \bar{X})^2 \frac{1}{N}$$

where $\bar{X}$ is substituted for $E(X)$ and $P(X = x_i) = 1/N$. While this is an intuitively plausible estimate of σ^2, it is not a good one in the sense to be defined in Chapter 9. Instead, the sample variance will be defined in this book as follows: the sample variance is the sum of the squared deviations divided by one less than the number of observations. Thus

$$S^2 = \sum_{i=1}^{N} (x_i - \bar{X})^2 \left(\frac{1}{N-1}\right)$$

The square root of S^2 is the sample standard deviation. Note that the analogy to the population variance still holds if $P(X = x_i) = 1/(N-1)$. When N is large, $\hat{S}^2$ and S^2 will be nearly equal, yet S^2 will be the only estimate of σ^2 used in this book. Older statistics text books tend to use $\hat{S}^2$ as the estimate of σ^2. Because of this, one should always find out how the sample variance is defined in any statistics text that one might use, especially if the book is somewhat old. The difference in definitions might make a difference in computations, especially if the sample sizes are small.

As was already shown, the sample average of the sample S: {3,6,8,9,14} is 8. Therefore the sample variance is given by

$$S^2 = \frac{(3-8)^2 + (6-8)^2 + (8-8)^2 + (9-8)^2 + (14-8)^2}{5-1}$$

$$= \frac{(-5)^2 + (-2)^2 + (0)^2 + (1)^2 + (6)^2}{4}$$

$$= \frac{25 + 4 + 0 + 1 + 36}{4} = \frac{66}{4} = 16.5$$

with the standard deviation given by $S = \sqrt{16.5} = 4.08$.

If the data have been grouped, then

$$S_x^2 = \sum_{k=1}^{K} \frac{f_k(x_k - \bar{X})^2}{N-1}$$

where x_k is the midpoint of the kth interval and f_k is the frequency in that interval. Generally, one does not use this formula to estimate σ_X^2 since the arithmetic involved becomes formidable under most circumstances. As with the determination of the sample average, a transformed scale simplifies the arithmetic. In addition, a formula analogous to the computing formula of a population variance helps reduce the arithmetic to an easily manageable form.

8-9 COMPUTING FORMULA FOR A SAMPLE VARIANCE ON A TRANSFORMED SCALE

Before deriving the computing formula for a sample variance, we state and prove the following theorem, which is based on Theorems 7-2 and 7-4.

Theorem 8-2

If $\mathrm{Var}(Y) = \sigma_Y^2$, and if $X = \beta Y + \alpha$, then $\mathrm{Var}(X)$ is given by

$$\mathrm{Var}(X) = \beta^2 \,\mathrm{Var}(Y)$$

Proof. By definition, $\mathrm{Var}(X) = \mathrm{Var}(\beta Y + \alpha)$. Since β and α are constants, it follows from Theorems 7-2 and 7-4 that

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(\beta Y)\\ &= \beta^2 \,\mathrm{Var}(Y)\end{aligned}$$

This completes the proof.

Since a sample variance is analogous to a population variance, it follows that if $X = \beta Y + \alpha$, then

$$S_x^2 = \beta^2 S_y^2$$

Thus, one need only compute S_y^2 and multiply it by β^2 and thereby obtain S_x.

Theorem 8-3

On the transformed scale for the midpoint values, the computing formula for the sample variance is given by

$$S_y^2 = \frac{N\left(\sum_{k=1}^{K} f_k y_k^2\right) - \left(\sum_{k=1}^{K} f_k y_k\right)^2}{N(N-1)}$$

Proof. By definition,

$$S_y^2 = \frac{1}{N-1}\sum_{k=1}^{K} f_k(y_k - \bar{Y})^2$$

If the binomial term is squared, and if the rules of summation of Section 4-4 are applied to the resulting sums, then the desired result follows. Thus

$$S_y^2 = \frac{1}{N-1}\sum_{k=1}^{K} f_k(y_k^2 - 2y_k\bar{Y} + \bar{Y}^2)$$

$$= \frac{1}{N-1}\left[\sum_{k=1}^{K} f_k y_k^2 - \sum_{k=1}^{K} 2f_k y_k \bar{Y} + \sum_{k=1}^{K} f_k \bar{Y}^2\right]$$

$$= \frac{1}{N-1}\left[\sum_{k=1}^{K} f_k y_k^2 - 2\bar{Y}\sum_{k=1}^{K} f_k y_k + \bar{Y}^2 \sum_{k=1}^{K} f_k\right]$$

$$= \frac{1}{N-1}\left[\sum_{k=1}^{K} f_k y_k^2 - 2\bar{Y}(N\bar{Y}) + \bar{Y}^2 N\right]$$

$$= \frac{1}{N-1}\left[\sum_{k=1}^{K} f_k y_k^2 - N\bar{Y}^2\right]$$

$$= \frac{1}{N-1}\left[\sum_{k=1}^{K} f_k y_k^2 - \frac{1}{N}\left(\sum_{k=1}^{K} f_k y_k\right)^2\right]$$

$$= \frac{N\left(\sum_{k=1}^{K} f_k y_k^2\right) - \left(\sum_{k=1}^{K} f_k y_k\right)^2}{N(N-1)}$$

This completes the proof.

To use this formula to compute S_x^2, one first computes S_y^2. For this, compute $\sum_{k=1}^{K} f_k y_k^2$ and $\sum_{k=1}^{K} f_k y_k$. The last of these, $\sum_{k=1}^{K} f_k y_k$, was computed when $\bar{X}$ was determined. As is recalled, $\sum_{k=1}^{K} f_k y_k = 17$. To simplify the determination of $\sum_{k=1}^{K} f_k y_k^2$, it is convenient to add column 6 to Table 8-7. Column 6 is found simply by multiplying column 4 times column 5. In this example, $\sum_{k-1}^{K} f_k y^2 = 105$. Thus

$$S_y^2 = \frac{N\left(\sum_{k=1}^{K} f_k y_k^2\right) - \left(\sum_{k=1}^{K} f_k y_k\right)^2}{N(N-1)}$$

$$= \frac{25(105) - (17)^2}{25(24)}$$

$$= \frac{2625 - 289}{600}$$

$$= \frac{2336}{600}$$

$$= 3.8933$$

and

$$\begin{aligned} S_x^2 &= \beta^2 S_y^2 \\ &= (5)^2 (3.8933) \\ &= 97.33 \end{aligned}$$

so that the sample standard deviation is given by

$$S_x = 9.87$$

The standard deviation of the population is about 10.

Finally, it should be noted that if the sample values are not grouped, then the computing formula for S_x^2 reduces to

$$S_x^2 = \frac{N\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i\right)^2}{N(N-1)}$$

For the sample X: {3,6,8,9,14},

$$\sum_{i=1}^{5} x_i = 3 + 6 + 8 + 9 + 14 = 40$$

$$\sum_{i=1}^{5} x_i^2 = 3^2 + 6^2 + 8^2 + 9^2 + 14^2 = 386$$

so that

$$S_x^2 = \frac{5(386) - (40)^2}{5(4)} = \frac{1930 - 1600}{20} = \frac{330}{20} = 16.5$$

as was shown earlier.

8-10 SAMPLE MEAN OR AVERAGE AND SAMPLE MEDIAN AS ESTIMATORS OF THE POPULATION CENTER

If a distribution is known to be symmetrical, the sample mean and the sample median are possible sample values that may be used to describe the population center. In most cases, the sample mean is preferred to the sample median as an estimate of the population center. Reasons for this preference will be explained by examples.

As a first example, consider seven boys and five girls observed playing on a school playground. After the play period, each of them was scored by five judges according to the amount of hostility shown toward other children during their play. The average hostility scores are shown in Table 8-8.

For the boys, $\bar{X}_B = 84.0$ and $\hat{M}_B = 72$. For the girls, $\bar{X}_G = 42.2$ and $\hat{M}_G = 40$. Using either measure of central tendency, one would conclude that the boys showed more hostility than the girls in this particular play period.

Table 8-8. Average hostility scores shown by 12 children playing, as determined by five independent judges.

Boys' names	*Boys' scores*	*Girls' names*	*Girls' scores*
Adam	83	Alice	74
Bob	71	Betty	40
Charles	72	Carol	48
Donald	63	Donna	26
Edward	111	Edna	23
Frank	67		
George	121		

Note that the median for the boys, 72, is based, for the most part, upon one observation only, namely the middle value, whereas the mean, 84.0, is based upon an equal weighting of all seven scores. The sample mean is said to utilize all the information that the sample has to offer about the center. The median uses less information, namely, the ordered ranking of the data. To clarify this point, consider the following three ordered samples:

Sample 1: $\{6, 8, 9, 12, 20, 44, 47\}$
Sample 2: $\{6, 8, 9, 12, 20, 44, 500\}$
Sample 3: $\{6, 8, 9, 12, 20, 44, 4593\}$

Each of these samples has the same median. However, their means are different. The means of the three samples are 20.9, 85.6, and 670.3.

On the other hand, these examples illustrate a minor weakness inherent in the sample mean as a measure of the population center—namely, that sample means are unduly affected by extreme values. Because of this tendency, sample averages are not recommended as suitable measures of the centers of probability distributions that are known to be *skewed*. Thus, economic and social statistics for skewed distributions are frequently reported in terms of medians rather than in terms of averages.

Suppose one wanted to combine the scores of the boys and those of the girls into one large sample and determine their combined average. This average is given by

$$\bar{X}_C = \frac{(83 + 71 + 72 + 63 + 111 + 67 + 121) + (74 + 40 + 48 + 26 + 23)}{7 + 5}$$

which can also be written as

$$\bar{X}_C = \frac{7(84.0) + 5(42.2)}{7 + 5} = 66.6$$

As this example suggests, the computing formula for a combined mean in the general case is given by

$$\bar{X}_C = \frac{N_B \bar{X}_B + N_G \bar{X}_G}{N_B + N_G}$$

This last equation shows that if the average value of two samples is known, the average of their combined sample is also known. One might think that this statement is also true of medians. However, it is not. To see this, consider the median of the combined sample. This median is given by

$$\hat{M}_C = 67$$

This value *cannot* be determined from

$$\hat{M}_C = \frac{N_B \hat{M}_B + N_G \hat{M}_G}{N_B + N_G}$$

since it gives an incorrect value for $\hat{M}_C$, as is shown:

$$\hat{M}_C = \frac{7(72) + 5(40)}{12} = 58.7$$

Thus, medians *cannot* be combined arithmetically, whereas means can. This suggests that sample medians are not very useful for complex analysis similar to that frequently encountered by the behavioral scientist. For the most part, medians tend to be dead-end statistics that cannot be used in statistical inference. Once they are known, they cannot be used for further analysis. The same is not true of sample averages.

Another reason why averages are preferred to sample medians as a measure of center is that sample means are more precise than sample medians. The truth and meaning of this statement are discussed in Chapter 9.

8-11 SUMMARY

One of the best procedures available to ensure a reasonable representation of a population is to select a random sample. A sample is said to be a simple random sample if

1. Every element of the sample is selected independently of all other elements of the population.
2. The probability of being included in the sample is equal for all elements in the population.

All standard statistical procedures are based upon the assumption of random sampling. If this kind of sampling is employed, one can obtain precise estimates of population parameters and one can test hypotheses concerning the numerical values that parameters assume.

If a universe is finite, random number tables may be used to ensure the selection of a random sample. If a universe is a hypothetical construct, such as eighth-grade students enrolled in a class in world geography, the most that one can do is hope that the sample selected is a random sample of that hypothetical population. Perhaps a wiser strategist might decide to redefine the target population and in its place use a finite population of eighth-grade world-geography students living in a specified school district. If this is done, then random number tables can be used to select a sample.

Once a random sample is selected, the data are generally organized into a frequency table from which histograms, relative-frequency polygons, or cumulative relative-frequency polygons are constructed. Frequency tables are also used to estimate population percentiles, expected values, and variances.

If the outcomes of a sample are denoted by X: $\{x_1, x_2, \ldots, x_N\}$, the sample mean, variance, and computing formula for the variance are defined by

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$S_x^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{X})^2}{N-1} = \frac{N\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i\right)^2}{N(N-1)}$$

If the data are summarized in a frequency table containing K mutually exclusive and exhaustive subclasses and if the midpoints are denoted by X: $\{x_1, x_2, \ldots, x_K\}$ with frequencies f: $\{f_1, f_2, \ldots, f_K\}$, then the sample mean, variance, and computing formula for the variance are given by

$$\bar{X} = \frac{1}{N} \sum_{k=1}^{K} f_k x_k$$

$$S_x^2 = \frac{\sum_{k=1}^{K} f_k (x_k - \bar{X})^2}{N-1} = \frac{N\left(\sum_{k=1}^{K} f_k x_k^2\right) - \left(\sum_{k=1}^{K} f_k x_k\right)^2}{N(N-1)}$$

If a code scale is used to transform the data with $Y = (X - \alpha)/\beta$, then one estimates $\bar{Y}$ and S_y^2 and then decodes back to the original scale with

$$\bar{X} = \alpha + \beta \bar{Y} \qquad \text{and} \qquad S_x^2 = \beta^2 S_y^2$$

In general, α is set equal to the midpoint of the interval with the greatest relative frequency, while β is set equal to the width of the intervals that must be equal in length for this method to have any sort of validity.

One of the arithmetic properties of a sample mean that has considerable use is that averages can be added, provided that they are weighted by their sample sizes.

Thus, if $\bar{X}_1$ and $\bar{X}_2$ are sample averages based on N_1 and N_2 observations, then the average of the sample formed by "pooling" the two samples is given by

$$\bar{X}_C = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2}$$

Sample medians do not have this property. In addition, sample medians do not utilize all of the information that the sample data contains concerning the center of a distribution. Finally, sample medians are not as precise as sample means when used as estimators of the expected value of a probability (usually symmetric) distribution. The full meaning of this last property is stressed in Chapter 9.

EXERCISES

***8-1.** Using Table A-5 of random numbers, select a random sample of 25 observations from Table 8-1. Using the class intervals of Table 8-2, prepare a frequency table and then construct the histogram, relative-frequency polygon, and cumulative relative-frequency polygon for the resulting data.

***8-2.** Estimate P_{25}, P_{50}, and P_{75} from the frequency table and from the cumulative relative-frequency polygon of Exercise 8-1.

***8-3.** Estimate $E(X)$ and σ for the data of Exercise 8-1.

***8-4.** In Table 3-1 it is stated that $P(50 \leqslant X \leqslant 59) = .34$. How does your sample compare with this population value? Assuming that your sample is from a normal population, compute $P(\bar{X} \leqslant X \leqslant \bar{X} + S)$. How does this compare with the population value? Do you think that a random sampling is a sampling procedure that has much to recommend its use? Explain.

***8-5.** In a study designed to measure the effects of reinforcement on learning, 25 kindergarten children were assigned to one of two experimental conditions on the basis of a coin tossing. If a head appeared, the student was assigned to the condition in which the correct response was rewarded with a piece of candy. Students assigned to the remaining condition were verbally rewarded by praise for a correct response. The dependent variable is the number of trials it took the student to make four correct responses in succession. The results of the coin tossing and the testing are as follows:

Candy reinforcement			*Verbal reinforcement*			
17	14	9	21	19	28	19
23	20	12	28	31	29	17
9	19	13	17	15	23	
6	10		16	22	24	

(a) What are the universes of this study? Are they real or abstract? Explain.
(b) To what populations can one make inferences?
(c) Estimate $E(X)$ and σ_X of each universe. What do the resulting statistics suggest about the parameters of the two reinforcement conditions of the study?

8-6. Sometimes sample means cannot be used to describe the center of a distribution since the form in which the data are collected makes arithmetic averaging impossible. This is frequently the case in questionnaire data or data reported by government agencies. The following data show the years of completed education in two extreme census tracts in Berkeley, California at the 1960 census. Note that the last category, 16 or more years, does not have an upper limit, and therefore no midpoint. For these data, estimate P_{10}, P_{50}, and P_{90}. What do the resulting statistics indicate about the years of completed education in these two census tracts?

Years of school completed	*Tract Be-0001-A*	*Tract Be-0006-A*
Persons 25 years old or over	2033	2684
No school years completed	87	13
Elementary:		
1–4 years	185	16
5–7 years	354	23
8 years	270	53
High school:		
1–3 years	427	144
4 years	415	452
College:		
1–3 years	178	469
4 years or more	117	1514

*8-7.** Often data are reported in intervals of unequal width. When this occurs, coding is not possible, even though midpoint assignments are possible. In a survey of the earning power of a junior college education following departure from school, the statistics reported in the following table were obtained. These results are based upon a sampling of 183 men

Yearly income in dollars	*Number of completed semesters of junior college*					*Total*
	ONE	TWO	THREE	FOUR	FIVE	
0–999	3	1	1	0	0	5
1,000–2,999	8	9	0	9	1	27
3,000–5,999	5	13	6	12	7	43
6,000–9,999	2	6	5	17	18	48
10,000–14,999	0	0	1	5	6	12
15,000–19,999	0	0	0	1	0	1
Total	18	29	13	44	32	136

who did not return to register at a certain junior college after completing the 1964 spring semester. The questionnaires were mailed three years following departure from this particular college.

(a) Estimate the expected incomes by number of completed semesters. What do these statistics suggest about the relationship that education has to earning power?
(b) Note that only 136 of the 183 men contacted responded to the survey. How do you think the lack of response might have biased the results? Explain.

8-8. Behavioral researchers frequently find that a small set of subjects does not perform according to the rest of the group and thereby tends to distort the statistics of an experimental condition. When this occurs, many behavioral researchers tend to discard data that are more than 4 sample standard deviations from the average, arguing that a score that is more than 3 standard deviations from expectation is, for all intents and purposes, unlikely With this in mind, consider the following set of data, which represents the number of trials it takes a college sophomore to learn a list of nonsense words backward:

$$X: \{17,23,19,16,8,12,27,63,17,38,21,32\}$$

(a) Estimate $E(X)$ and σ_X with, and without, the unusual score of 63. What is the effect on the estimate if this unusual score is included or excluded from the analysis?
(b) Frank E. Grubbs (1969) presented Table A-6, which can be used to test whether an "outlier" can be rejected as belonging to a sample with a probability of a type I error controlled at $\alpha = .05$, .025, or .01. For his test one computes

$$T = \frac{x_N - \bar{X}}{S} \quad \text{or} \quad T = \frac{\bar{X} - x_1}{S}$$

where $\bar{X}$ = mean of the complete sample, S = standard deviation of the complete sample, and x_N = largest suspected observation or x_1 = smallest suspected observation. If T is larger than the tabled value, the suspected outlier is rejected. Use Grubbs' test to determine if $x_{12} = 63$ is an outlier at $\alpha = .05$.
(c) What are your feelings about discarding unusual data? What would you do in this case? Defend your position.

***8-9.** Ordered qualitative variables are frequently quantified by superimposing upon the classes a Likert scale. An example of such a scale and its numerical scoring is: {strongly agree, moderately agree, moderately disagree, strongly disagree} and {1,2,3,4}. Assign this scale to the statistics of Exercise 3-10. Estimate the expected Likert scores for the two marginal distributions of the table. Also estimate the standard deviations of the two distributions. What do the resulting statistics suggest about the public's attitudes toward the integration of elementary and junior high schools?

***8-10.** In a study designed to measure the ability of individuals in distinguishing among 4 well-known cola drinks, 48 college sophomores were given a blindfold test in which 4 cola drinks were presented in random order. The dependent variable of the study is the

number of colas that one student can correctly identify. The results were as shown in the following table.

Number of correct identifications	*Frequency*
0	20
1	20
2	5
3	0
4	3
Total	48

(*a*) Estimate $E(X)$ and σ_X for these data.

(*b*) How do these estimates compare with the theoretical values of Table 6-3, which is the mathematical model of this study for which it is assumed that a correct matching is a chance event?

9 SAMPLING DISTRIBUTIONS OF STATISTICAL ESTIMATORS

Since it is obviously impractical to poll the nation on anything less important than the selection of a President, one cherished statistical tool is the sample. Not even statisticians can agree on how big or good a sample can be relied upon as representing the whole Dr. Alfred C. Kinsey's celebrated reports were criticized by statisticians not so much for their moral implications but because they made sweeping presumptions on the basis of too small a sample (in the male study, only 5,300 men provided data) The Nielsen ratings, by which television programs live or die, have been justly attacked because Nielsen recorders are necessarily hooked to the sets of those viewers willing to have a recorder—a special class by definition, whose tastes may or may not correspond with those of the unpolled millions of the total TV audience.

9-1 PROPERTIES OF POINT ESTIMATORS

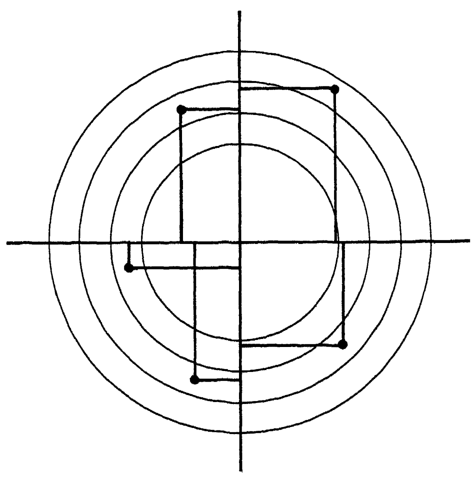
A point estimate of a parameter of a probability distribution is a unique number that may be computed from a random sample selected from a population. In general, it is the estimate that intuition would suggest be used. A point estimate is to be contrasted with an interval estimator, which will be defined in Chapter 11. Upon completion of the discussion on interval estimates, the distinction between point and interval estimates will become clear. One need not dwell on the distinction too much at this time

Instead, imagine two marksmen shooting at the center of a target, each trying to hit the bull's-eye. Suppose marksman A takes five shots at the target and makes the configuration of shots shown in Figure 9-1. Even though all shots of marksman A are on target, most marksmen would not be impressed with his shot group. While it is true that the sum of the vertical and the sum of the horizontal deviations from the bull's-eye is 0, it is easy to see that on repeated shooting, marksman A tends to show a large variance or fluctuation about the center of the bull's-eye. A marksman who can consistently shoot around the bull's-eye is said to be an accurate shot.

Consider marksman B, who makes the configuration of shots shown in Figure 9-2. For this marksman the vertical deviations and horizontal deviations of his shot group do not sum to 0. On the average, marksman B does not fluctuate about the center of the bull's-eye but consistently shoots about a "bull's-eye" in the upper right quadrant of the target. While he may not be very accurate, he is certainly precise. He would be called a *precision* shooter, who could become *accurate* by adjusting his sight.

Consider a behavioral scientist who has two methods for estimating the center of a probability distribution. Certainly he wants to use the method that is accurate

Figure 9-1. Shot group for an accurate but nonprecise shooter.



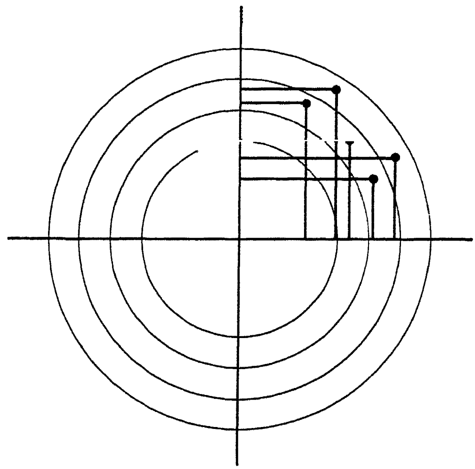


Figure 9-2. Shot group for an inaccurate but precise shooter.

so that it produces correct values on the average, even though it may miss the correct value on each trial. Furthermore, he would like the estimating procedure to be precise so that upon repeated trials it would give near identical results over and over again. In this sense he desires an estimation method that is accurate like marksman A and that is precise like marksman B.

As these two examples suggest, a sharpshooter is a person who is an accurate and precise shooter. In an analogous manner, one could look at the bull's-eye as a parameter to be estimated, such as $E(X)$. The marksmen can be equated to two different estimation procedures that could be used for estimating $E(X)$, such as the sample mean and the sample median. On repeated samplings, they produce estimators that fluctuate about $E(X)$. The method that has the greatest accuracy and precision is clearly the optimum method to use. As might be suspected, this optimum estimation method is embodied in the sample mean. The truth of this statement will be illustrated in this chapter.

Consider the two most commonly used estimators of $E(X)$ for a symmetrical probability distribution, $\bar{X}$ and $\hat{M}$. If there is a choice, use the one that tends to give the correct value on the average and that tends to give the same value or nearly the same value on repeated trials. As one might suspect, the sample mean and sample median are equally accurate, provided that the underlying distribution is symmetrical. However, the sample mean is more precise and for that reason it is preferred. The meaning of these statements will be clarified in the next sections.

9-2 THE SAMPLING DISTRIBUTION OF $\bar{X}$ AND $\hat{M}$

Consider the population of 500 students of Table 8-1 who were given the test with unknown expected value and unknown variance. Consider all possible samples

of size 5 that can be drawn without replacement. The total number of such samples is given by

$$T = \binom{500}{5} = \frac{(500)(499)(498)(497)(496)}{(5)(4)(3)(2)(1)} = 255{,}244{,}687{,}600$$

Each of these 255,244,687,600 samples consists of one point in the sample space defined by the cartesian product of the numbers in Table 8-1, but in which repeated appearance of the same elements is not permitted. For example, $(x_1,x_2,x_3,x_4,x_5) \in S$, while $(x_1,x_1,x_2,x_3,x_4) \notin S$. Thus, the total set of possible samples of size 5 that one may generate without replacement from the complete set of 500 numbers is given by S: {(55,51,65,60,64), (55,51,65,60,59), . . . , (49,38,61,41,62)}.

Consider the sample average $\bar{X}$ for each of these 255,244,687,600 samples. These averages are given by $\bar{X}$: {59.0, 58.0, . . , 50.2}. The totality of these outcomes gives rise to a discrete variable $\bar{X}$ (the sample average) that has a probability distribution, an expected value, and a variance. While it is physically impossible to determine the distribution exactly, one can estimate it by drawing a random sample of averages and studying the observed distribution to make inferences about the population of averages. Studies of this nature are called Monte Carlo studies since they correspond to an empirical testing of a roulette wheel. At Monte Carlo, roulette wheels contain 37 numbers: {0,1,2, . . . ,36}. One way to test the balance of the wheel is to spin it 3,700 times. If the wheel is balanced, one should observe about 100 instances of each number (0,36), since the distribution of X is known to be $U(0,36)$. Many times, the exact distribution of X is unknown. One way to determine it is to empirically

Table 9-1. Frequency table of the sample averages generated by 32 students taking a beginning course in statistics.

True intervals	*Midpoint* x_k	*Frequency* f_k	*Relative frequency*	*Code scale* y_k	$f_k y_k$	$f_k y_k^2$
36.5–39.5	38	1	.003	−4	−4	16
39.5–42.5	41	18	.056	−3	−54	162
42.5–45.5	44	34	106	−2	−68	136
45.5–48.5	47	68	.212	−1	−68	68
48.5–51.5	50	92	.288	0	0	0
51.5–54.5	53	69	.216	1	69	69
54 5–57 5	56	23	.072	2	46	92
57.5–60 5	59	11	.034	3	33	99
60.5–63.5	62	4	013	4	16	64
		320	1.000		−30	706

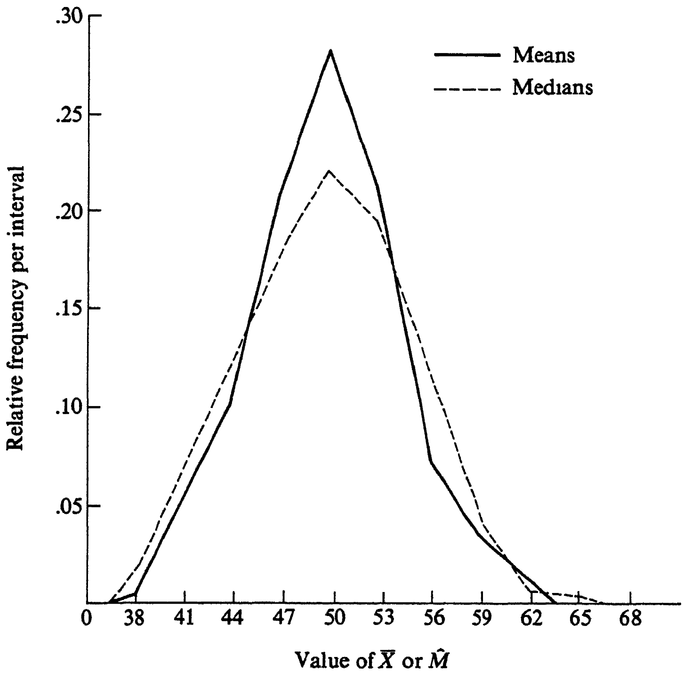
generate it by a large number of identical trials. When such an empirical study is performed, one says that a Monte Carlo investigation has been made.

For this study, each of 32 students in a beginning statistics class selected 10 samples of size 5 from the population of Table 8-1. The average of each sample was computed. The frequency distribution of this sampling experiment is shown in Table 9-1 and Figure 9-3.

Examination of the data of Table 9-1 suggests that the *sampling distribution* of $\bar{X}$ is symmetric. If, in fact, one had the entire distribution of 255,244,687,600 averages, the symmetry would be pronounced and its closeness to the normal distribution would be striking. This is not unexpected because of the central limit theorem. As is recalled, $T = X_1 + X_2 + \cdots + X_N$ tends to a normal distribution with $E(T) = N\mu$ and $\text{Var}(T) = N\sigma^2$, provided that the X's are identically and independently distributed.

The variable $\bar{X} = \frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5) = \frac{1}{5}T$ is of this form and its distribution should be close to normal. Since $\bar{X}$ is a constant multiple of T, $\bar{X}$ has the same distribution as T except that it has a different scale and center. The parameters of the distribution of $\bar{X}$ are stated in the following theorem.

Figure 9-3. Empirical sampling distribution of the 320 sample means and medians for samples of size 5 from a population with $E(X) = 50$ and $\sigma_X = 10$.



Theorem 9-1

The parameters of the sampling distribution of $\bar{X}$ for samples of size N are given by

$$E(\bar{X}) = \mu \qquad \text{and} \qquad \text{Var}(\bar{X}) = \frac{\sigma^2}{N}$$

Proof. According to Theorem 7-3,

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{N}T\right) \\ &= \frac{1}{N}E(T) \\ &= \frac{1}{N}N\mu \\ &= \mu \end{aligned}$$

According to Theorem 7-4,

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{N}T\right) \\ &= \left(\frac{1}{N}\right)^2 \text{Var}(T) \\ &= \left(\frac{1}{N}\right)^2 N\sigma^2 \\ &= \frac{\sigma^2}{N} \end{aligned}$$

This completes the proof.

For the sampling experiment the *average of the averages* is given by

$$\begin{aligned} \bar{X}_X &= \beta \bar{Y} + \alpha \\ &= 3\left(\frac{-30}{320}\right) + 50 \\ &= -.3 + 50 \\ &= 49.7 \end{aligned}$$

This is very close to the expected value, which in this case is exactly equal to 50. If all possible samples had been available for this analysis, it would have been found that the average of all the possible averages would have been equal to 50.

In addition to the determination of $\bar{X}$, the 32 students of the beginning statistics class were asked to determine the sample medians. The empirical distribution of

the sample medians is shown in Table 9-2 and Figure 9-3. Again it is seen that the distribution of the medians is symmetrical and that the *average of the medians* is given by

$$\begin{aligned}\bar{X}_{\hat{M}} &= \beta\bar{Y}+\alpha\\ &= 3\left(\frac{-21}{320}\right)+50\\ &= -.2+50\\ &= 49.8\end{aligned}$$

If all possible samples had been available, this value would have been exactly equal to $E(X)$.

Thus, it has been empirically shown that the sample median and the sample average are equally accurate estimators of $E(X)$. Figuratively speaking, both estimators fluctuate about the bull's-eye of $\mu = 50$. In terms of the language of the statistician, it is said that $\bar{X}$ and $\hat{M}$ are *unbiased* estimators of $E(X)$. The meaning of this statement is shown in the following section. Finally, it should be noted that if the distribution of X is skewed, $\bar{X}$ continues to provide an unbiased estimator of μ while $\hat{M}$ provides a biased estimator.

Table 9-2. Frequency table of the sample medians generated by 32 students taking a beginning course in statistics.

True intervals	*Midpoint* x_k	*Frequency* f_k	*Relative frequency*	*Code scale* y_k	$f_k y_k$	$f_k y_k^2$
36.5–39 5	38	8	.025	−4	−32	128
39.5–42 5	41	23	.072	−3	−69	198
42.5–45.5	44	33	.103	−2	−66	132
45.5–48.5	47	61	.190	−1	−61	61
48.5–51.5	50	71	.222	0	0	0
51.5–54.5	53	67	.209	1	67	67
54.5–57.5	56	39	.122	2	78	156
57 5–60 5	59	13	.041	3	39	117
60.5–63.5	62	2	.006	4	8	32
63.5–66.5	65	3	.009	5	15	75
		320	.999		−21	966

9-3 UNBIASED ESTIMATORS (COMPARISON OF $\bar{X}$ AND $\hat{M}$)

Let $\hat{\theta}$ be a formula for estimating a parameter θ of a population. Upon repeated sampling, $\hat{\theta}$ has a probability distribution with expected value $E(\hat{\theta})$ and variance

$\sigma_{\hat{\theta}}^2$. If the expected value of the sampling distribution of $\hat{\theta}$ is given by $E(\hat{\theta}) = \theta$, then the estimating procedure defined by $\hat{\theta}$ is an unbiased estimating procedure and the estimator $\hat{\theta}$ is said to be an unbiased estimator of θ. If $E(\hat{\theta}) = \theta + \beta$, then $\hat{\theta}$ is said to be a biased estimator of θ and the amount of the bias is equal to β.

As will now be shown, $\bar{X}$ (the sample average) is always an unbiased estimator for $E(X)$ or μ (the population expected value).

Theorem 9-2

The sample mean is an unbiased estimator of μ, regardless of the form of the probability distribution of X. Mathematically, if

$$\hat{\theta} = \bar{X} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

then

$$E(\bar{X}) = \mu$$

Proof By definition,

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \cdots + X_N}{N}\right)$$

Since N is a constant, it may be removed from the expectation symbol. Thus

$$E(\bar{X}) = \frac{1}{N}[E(X_1 + X_2 + \cdots + X_N)]$$

By Theorem 6-1, the expected value of a sum is the sum of the expected values. Thus

$$E(\bar{X}) = \frac{1}{N}[E(X_1) + E(X_2) + \cdots + E(X_N)]$$

Since all the X's have identical distributions,

$$E(\bar{X}) = \frac{1}{N}(\mu + \mu + \cdots + \mu)$$

$$= \frac{1}{N}N\mu$$

$$= \mu$$

This completes the proof.

A corresponding theorem for $\hat{M}$ is stated without proof.

Theorem 9-3

If the probability distribution of X is symmetrical, $\hat{M}$ is also an unbiased estimator of μ.

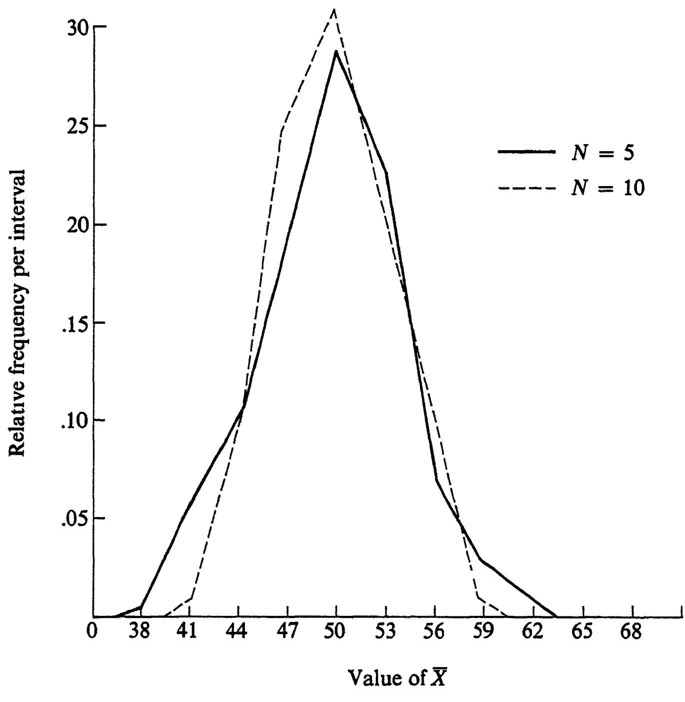
9-4 VARIANCES OF SAMPLING DISTRIBUTIONS

For the Monte Carlo study of Section 9-2, each student was asked to select five random samples each of size 10 and to compute their means and medians. The corresponding distributions are shown in Figures 9-4 and 9-5.

As can be seen, the distributions are symmetrical and much more compact about the expected values than those of samples of size 5. This compactness is summarized in Table 9-3. Notice that $S^2_{\bar{X}_5} > S^2_{\bar{X}_{10}}$ and that $S^2_{\hat{M}_5} > S^2_{\hat{M}_{10}}$. Thus, the variability in averages for samples of size 10 is less than the variability of samples of size 5. The same statement is also true for the median. On this basis, one would conclude that samples of size 10 are more precise than samples of size 5. However, what is of more interest is that $S^2_{\bar{X}_5} < S^2_{\hat{M}_5}$ and $S^2_{\bar{X}_{10}} < S^2_{\hat{M}_{10}}$. Thus, sample averages are more precise than sample medians based on the same sample size. As a statistician would say, a sample average is more efficient as an estimator of the population expected value than is a sample median.

Thus, while one could not choose between $\bar{X}$ and $\hat{M}$ on the basis of their accuracy or unbiasedness, one can make a decision as to which is the better measure on the basis of their precision or efficiency. The estimate with greater precision is said to be an *efficient estimator.*

Figure 9-4. Empirical sampling distributions of the sample mean for samples of size 5 and 10 from a population with $E(X) = 50$ and $\sigma_X = 10$



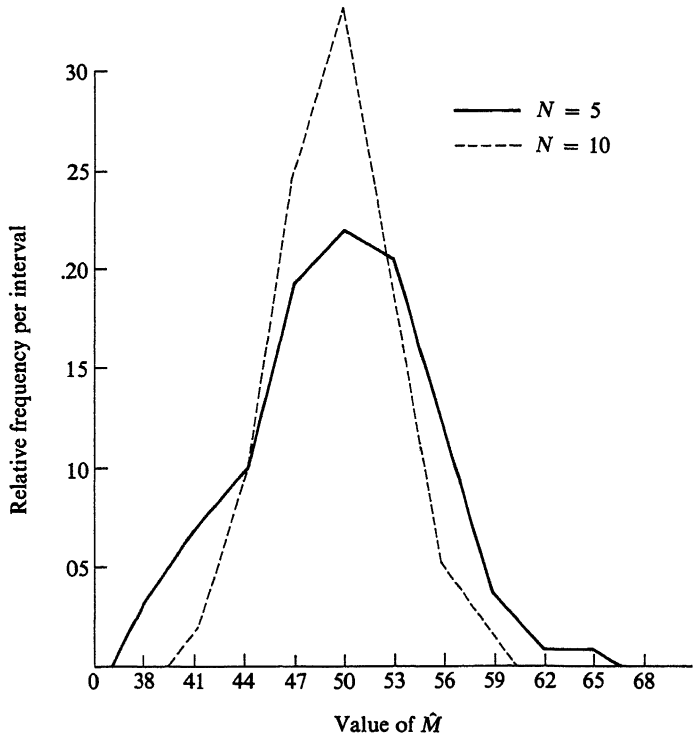


Figure 9-5. Empirical sampling distributions of the sample median for samples of size 5 and 10 from a population with $E(X) = 50$ and $\sigma_X = 10$

Table 9-3. Table of sample variances for the four Monte Carlo distributions generated by the students in a beginning course in statistics.

Statistic	*Sample size* 5	10
Sample mean $\bar{X}$	19.84	10.64
Sample median $\hat{M}$	27.22	14.16

9-5 EFFICIENT ESTIMATORS

A statistic $\hat{\theta}_1$ is said to be a more efficient estimator of an unknown parameter θ than another statistic $\hat{\theta}_2$ if the variances of their sampling distributions are such that the $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$.

Mathematical statisticians define efficiency in a slightly different form. What is called efficiency in this book would normally be called relative efficiency by mathematical statisticians and would be defined in terms of expected mean squares, a concept that has not yet been introduced.

As might be suspected, there are some statistical measures that are more accurate than others, but less precise. When this occurs there is a definite tendency to choose the efficient estimator. However, this is not always the case, as will be seen for a very important statistical measure. Efficiency, as defined above, indicates that it is a relative concept. An estimator is or is not efficient when compared to another. This suggests the following definition of relative efficiency. If $\hat{\theta}_1$ is the efficient estimator of a parameter and if θ_2 is any other statistic that can be used as an estimator of a parameter, the relative efficiency of $\hat{\theta}_2$ compared to $\hat{\theta}_1$ is measured by

$$r_e = \frac{\mathrm{Var}(\hat{\theta}_1)}{\mathrm{Var}(\hat{\theta}_2)}$$

From the Monte Carlo study, the estimated relative efficiency of the median compared to the mean for samples of size 5 is given by

$$\hat{r}_e = \frac{S^2_{\bar{X}_5}}{S^2_{\hat{M}_5}} = \frac{19.84}{27.22} = .73$$

and for samples of size 10 is given by

$$\hat{r}_e = \frac{S^2_{\bar{X}_{10}}}{S^2_{\hat{M}_{10}}} = \frac{10.64}{14.16} = .75$$

Thus, the sample median is only 75 percent as efficient as the sample average. The median compared to the mean is said to have an efficiency of 75 percent.

For the normal distribution, the relative efficiency of the median compared to the mean is, for samples of size 5, given by $r_e = .697$. For samples of size 10, $r_e = .723$, and for exceptionally large samples,

$$r_e = \frac{\mathrm{Var}(\bar{X})}{\mathrm{Var}(\hat{M})} = \frac{2}{\pi} = 63.7 \text{ percent}$$

As suggested by these figures, the efficiency of the median when compared to the mean varies in a nonmonotonic way as N increases. When N is small, the efficiency is quite high. The efficiencies for small samples are shown in Table 9-4. For the most part, a sample median based on 100 observations is as precise an estimator of $E(X)$ as is a sample average based on 64 observations. Although the mean and median are unbiased estimators of the expected value of the normal distribution, the efficiency of the mean is better (i.e., the mean is closer to μ most of the time).

Table 9-4. Efficiency of the median when compared to the mean for X normal.

N	*Efficiency*
2	1.000
3	743
4	.838
5	.697
10	723
15	.656
20	681
∞	.637

9-6 THE VARIANCE OF THE SAMPLING DISTRIBUTION OF $\bar{X}$

While it has not been stated, the variance of the population of Table 8-1 is equal to 100. As noted, $S^2_{\bar{X}_5} = 19.84$ and $S^2_{\bar{X}_{10}} = 10.64$. As these results suggest,

$S^2_{\bar{X}_5} = 19.84$ is approximately equivalent to $\frac{1}{5}(100)$

and

$S^2_{\bar{X}_{10}} = 10.64$ is approximately equivalent to $\frac{1}{10}(100)$

These results also suggest that in the population of $\bar{X}$'s,

$$\sigma^2_{\bar{X}} = \frac{1}{N}\sigma^2_X$$

and, indeed, this is the case. While this was derived in Theorem 9-1, it is so important in statistical theory and methodology that its derivation is repeated again, but with a different proof.

Theorem 9-4

The variance of the sampling distribution of $\bar{X}$ is given by

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{N} \qquad \text{or} \qquad \sigma^2_{\bar{X}} = \frac{\sigma^2_X}{N}$$

Proof. By definition,

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{X_1 + X_2 + \cdots + X_N}{N}\right)$$

By Theorem 8-2,

$$\mathrm{Var}(\bar{X}) = \frac{1}{N^2}\mathrm{Var}(X_1 + X_2 + \cdot \cdot + X_N)$$

Since the X's are statistically independent,

$$\mathrm{Var}(\bar{X}) = \frac{1}{N^2}[\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_N)]$$

Since the X's have identical distributions,

$$\mathrm{Var}(\bar{X}) = \frac{1}{N^2}[\sigma_X^2 + \sigma_X^2 + \cdots + \sigma_X^2] = \frac{1}{N}\sigma_X^2$$

This completes the proof.

As the sample size N gets larger, the $\mathrm{Var}(\bar{X})$ gets smaller. This suggests that a mean based upon a sample of size 10 has greater precision than a mean based upon a sample of size 5. If one were interested only in accuracy, a sample size of 10 would be as good as a sample size of 5 or any other size because, as was seen, their expected values are the same. The expected value of the sample mean is the same as the mean of the population regardless of how large the sample is. However, precision does not remain constant as sample sizes increase. As the sample exhausts the population, $\mathrm{Var}(\bar{X}) = \mathrm{Var}(X)/N$ approaches 0. This is why large samples are desirable. They tend to be more precise. For the Monte Carlo empirical investigation,

$$r_e = \frac{S_{\bar{X}_5}^2}{S_{\bar{X}_{10}}^2} = \frac{10.64}{19.84} = .54$$

so that a sample of size 5 has about half the efficiency of a sample of size 10 for estimating the expected value of a probability distribution.

9-7 THE THEORETICAL SAMPLING DISTRIBUTION OF MEANS AND MEDIANS

The exact sampling distributions of $\bar{X}$ and $\hat{M}$ are normal in form when the parent population is normal. When the parent population is not normal, the sampling distribution of $\bar{X}$ can be approximated by the normal distribution, provided that the sample size is sufficiently large. Unfortunately, these results cannot be proved with the materials developed in this book. However, an attempt is made in this section to show that the theorem is true for the special population of 500 reading scores.

According to the central limit theorem, the theoretical sampling distribution of $\bar{X}$ is approximately normal with $E(\bar{X}) = \mu$ and $\mathrm{Var}(\bar{X}) = \sigma^2/N$. For samples of size 5,

$$E(\bar{X}) = E(X) = 50$$

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{N} = \frac{100}{5} = 20$$

$$\sigma_{\bar{X}} = 4.48$$

This theoretical distribution of $\bar{X}$ is shown in Figure 9-6. Its similarity to the empirical distribution of Figure 9-4 should be noted. For samples of size 10,

$$E(\bar{X}) = E(X) = 50$$

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{N} = \frac{100}{10} = 10$$

$$\sigma_{\bar{X}} = 3.16$$

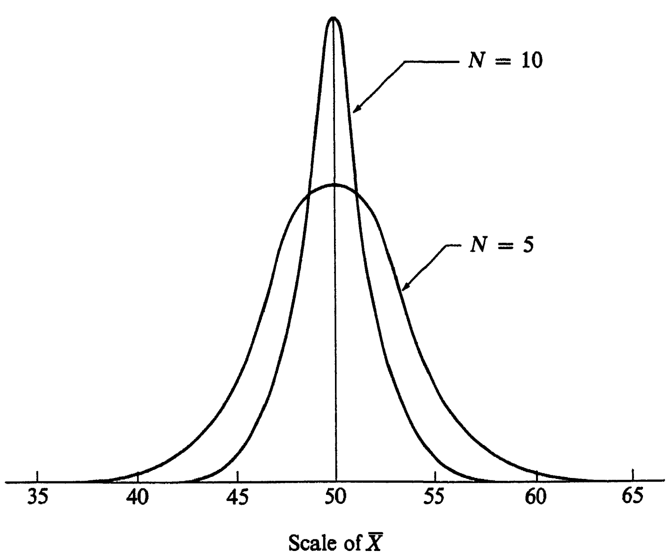
This distribution of $\bar{X}$ is also shown in Figure 9-6. Again, this distribution should be compared to the distribution shown in Figure 9-4. For samples of size 10, 95 percent of the sample means are between 43.68 and 56.32. For samples of size 5, the corresponding 95 percent range is 41.04 to 58.96. The larger interval for samples of size 5 reflects the greater precision of samples of size 10 over samples of size 5.

Theorem 9-5

The variance of the sampling distribution of $\hat{M}$ for large samples, provided that X is normal, is given by

$$\sigma_{\hat{M}}^2 = 1.57 \frac{\sigma^2}{N}$$

Figure 9-6. Distribution of $\bar{X}$ for samples of size 5 and 10 for which X is approximately normal with $E(X) = 50$ and $\sigma_X = 10$.



Proof. Consider the sampling distribution of the medians. The relative efficiency of $\hat{M}$ to $\bar{X}$ is defined as

$$r_e = \frac{\mathrm{Var}(\bar{X})}{\mathrm{Var}(\hat{M})}$$

Therefore,

$$\mathrm{Var}(\hat{M}) = \frac{1}{r_e}\mathrm{Var}(\bar{X})$$

However,

$$\mathrm{Var}(\bar{X}) = \frac{\sigma_X^2}{N}$$

and for large samples from normal distributions,

$$r_e = \frac{2}{\pi}$$

Therefore,

$$\mathrm{Var}(\hat{M}) = \frac{\pi}{2}\frac{\sigma_X^2}{N} = 1.57\frac{\sigma_X^2}{N}$$

This completes the proof.

The standard deviation of the medians is given by

$$\begin{aligned}\sigma_{\hat{M}} &= \sqrt{\frac{1.57}{N}}\,\sigma_X \\ &= \frac{1.25\sigma_X}{\sqrt{N}} \\ &= \frac{5}{4}\frac{\sigma_X}{\sqrt{N}}\end{aligned}$$

For the samples of size 5,

$$\sigma_{\bar{X}} = \frac{10}{\sqrt{5}} = 4.4722$$

$$\begin{aligned}\sigma_{\hat{M}} &= (1.25)(4.4722) \\ &= 5.59\end{aligned}$$

which is very close to the sample value

$$S_{\hat{M}_5} = 5.22$$

For samples of size 10,

$$\sigma_{\hat{M}} = (1.25)\frac{10}{\sqrt{10}} = 3.95$$

which is also close to the sample value of

$$S_{\hat{M}_{10}} = 3.77$$

The sampling distribution of $\hat{M}$ for samples of size 5 is shown in Figure 9-7. The empirical distribution for this variable is shown in Figure 9-5. The similarity between theory and actual results is quite striking. Thus, 95 percent of the sample medians are between 38.8 and 61.2. For samples of size 10, the corresponding interval is 42.1 to 57.9.

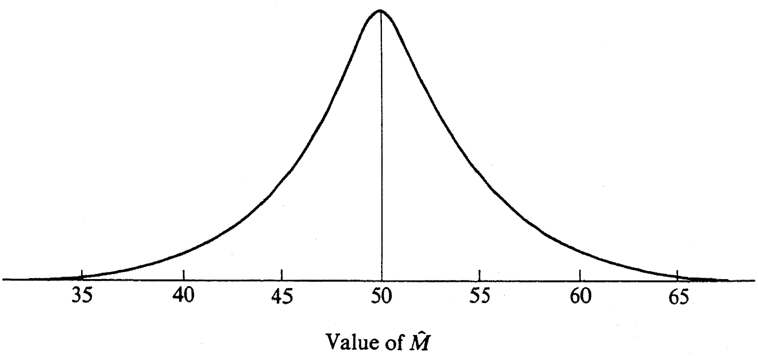
9-8 USES OF SAMPLING DISTRIBUTIONS

The sampling distribution of a statistic is of major importance in all of the methods of statistical inference discussed in this book. The methods to be developed constitute the major contribution of statistics to empirical research. This importance can be hinted at by the following example.

Consider a population of students who are to be sampled and then given a Stanford-Binet intelligence test. Suppose that the population of direct interest consists of students from upper-social-class areas. Past experience suggests that children with this background tend to have above-average IQs. As is well known, the expected IQ for the Stanford-Binet in the general population is 100 and the standard deviation is 16.

Suppose a random sample of 25 students is to be selected from the general population. If one could determine the largest and smallest averages that one should expect from that population, one could compare the observed outcome to the expected range of outcomes. If the observed outcome can be easily explained as a

Figure 9-7. Distribution of $\hat{M}$ for samples of size 5 for which X is approximately normal with $E(X) = 50$ and $\sigma_X = 10$.



possible member of the set of expected averages, then one would conclude that the IQ for the population of students from which the sample came also has $\mu = 100$. Otherwise, one would conclude the opposite.

According to the central limit theorem, the sampling distribution of $\bar{X}$ tends to be normal with

$$E(\bar{X}) = E(X) = 100$$

and

$$\text{Var}(\bar{X}) = \frac{\sigma_X^2}{N} = \frac{16^2}{25}$$

and

$$\sigma_{\bar{X}} = \tfrac{16}{5} = 3.2$$

Since the probability distribution is normal, almost all of the sample means are in the interval from

$$\bar{X}: \{E(\bar{X}) - 3\sigma_{\bar{X}} < \bar{X} < E(\bar{X}) + 3\sigma_{\bar{X}}\}$$
$$\bar{X}: \{100 - 3(3.2) < \bar{X} < 100 + 3(3.2)\}$$
$$\bar{X}: \{90.4 < \bar{X} < 109.6\}$$

Thus, one can feel quite confident that no sample will be selected from the general population that does not have its average in this particular range.

Suppose now that the sample average for the 25 children actually selected for the testing is given by $\bar{X} = 108.7$. On the surface, it appears that this value could easily be an element from the sampling distribution of $\bar{X}$ since it is well within 3 standard deviations of the expected average.

Suppose that another researcher in another community sampled essentially the same population but used a sample of 100 students. Suppose also that his sample average was equal to 108.7. Would his decision be the same? For this second case,

$$E(\bar{X}) = E(X) = 100$$

and

$$\text{Var}(\bar{X}) = \frac{\sigma_X^2}{N} = \frac{16^2}{100}$$

and

$$\sigma_{\bar{X}} = 1.6$$

Since sample means have a normal distribution, the $\pm 3\sigma_{\bar{X}}$ central range for $\bar{X}$ is given by

$$\bar{X}: \{E(\bar{X}) - 3\sigma_{\bar{X}} < \bar{X} < E(\bar{X}) + 3\sigma_{\bar{X}}\}$$
$$\bar{X}: \{100 - 3(1.6) < \bar{X} < 100 + 3(1.6)\}$$
$$\bar{X}: \{95.2 < \bar{X} < 104.8\}$$

Clearly, this researcher would not draw the same conclusion. $\bar{X} = 108.7$ is not a likely element of $N(100, 1.6^2)$. Thus, he might conclude that children in this population tend to have above-average IQ.

It is worth noting that the sampling distribution of means is a hypothetical construct; it is a pure abstraction that has significant utility for decision making. As an abstraction, it is known that its expected value is exactly equal to the mean of the population. The standard deviation of this abstraction is intimately connected to the standard deviation of the original population that gives rise to this hypothetical construct. In fact, the standard deviation of the distribution of means is equal to the standard deviation of the original or parent population divided by the square root of N:

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{N}}$$

Furthermore, the distribution of $\bar{X}$ can be adequately described by a normal distribution with parameters $E(X)$ and $\sigma_{\bar{X}}$.

In any one study, the sampling distribution of means is not explicitly determined because the financial costs would be prohibitive. However, it is *always* in the background and is *always* invoked in explaining statistical significance. The importance of sampling distributions *cannot* be overemphasized. Unless a true understanding is had of this major statistical concept, most of what follows will not make sense. Therefore, one is advised to read and reread the material of this chapter until the significance of the discussion is understood. The material presented in this chapter is, perhaps, the most difficult of all the material presented in a beginning statistics course. However, this material serves as the basic building block on which the remaining structures of statistical theory and methodology are constructed. Again, one is advised to learn this material well.

Finally, one might wonder if the theory of sampling distributions can be applied to the analysis of the Kinsey data or the Nielsen television ratings of the *Time* magazine excerpt. Unfortunately, they cannot. It is probably worth mentioning that the interviewed subjects of the Kinsey report were volunteers. That is, they willingly agreed to tell Kinsey about their personal sexual experiences. Whether their responses represent the general population of males may or may not be true. In addition, one is probably quite justified in questioning the validity of the results of the Neilsen television ratings. People who are willing to have a recorder hooked to their TV set may indeed represent a special class of TV viewers whose tastes may or may not correspond to those of the unpolled millions of the total TV audience. In any case, the application of the theory of sampling distributions to behavioral data is dependent upon random sampling.

9-9 SAMPLING DISTRIBUTION OF VARIANCE MEASURES

Just as measures of central tendency (mean and median) have sampling distributions,

measures of variation also have sampling distributions. As might be suspected, *all* measures based upon sample values have sampling distributions.

Just as there are many ways to estimate the center of a distribution, there are many ways to measure the variance. The two most frequently encountered methods are summarized in the following two formulas:

$$S^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N-1}$$

and

$$\hat{S}^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}$$

The first formula should be immediately recognized as the preferred formula for sample variance and the second formula as the intuitive estimate, which has been rejected. The reasons for this preference and rejection will now be shown.

As defined earlier, an estimation procedure is said to be unbiased if the expected value of the sampling distribution of the statistic generated by the procedure is equal to the value being estimated.

For the 320 samples generated by the students of the author's class, the following results were obtained:

$$\overline{S^2} = 99.96 \qquad \text{and} \qquad \overline{\hat{S}^2} = 79.97$$

Since it is known in this case that $\sigma_X^2 = 100$, it appears that $\hat{S}^2$ tends to be too low and on the average misses the true value by about 20 points. S^2 on the average is very close to σ_X^2. If all 255,244,687,600 samples had been obtained, it would be seen that

$$E(S^2) = \sigma^2$$

and

$$E(\hat{S}^2) = \frac{N-1}{N}\sigma^2 = \sigma^2 - \frac{\sigma^2}{N}$$

Thus, it follows that S^2 is an unbiased estimator of σ^2 and that $\hat{S}^2$ is biased. In fact, the amount of the bias in $\hat{S}^2$ is given by $\beta = -\sigma^2/N$, so that on the average it tends to be too small. It is mainly for this reason that S^2 has been preferred over $\hat{S}^2$.

Even though S^2 is unbiased, it is not as efficient as $\hat{S}^2$. To see this, let $\mathrm{Var}(S^2) = \sigma_{S^2}^2$.

Also note

$$\hat{S}^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}$$

$$= \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N} \frac{N-1}{N-1}$$

$$= \frac{N-1}{N} \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N-1}$$

$$= \frac{N-1}{N} S^2$$

Employing Theorem 7-4, we see that

$$\operatorname{Var}(\hat{S}^2) = \operatorname{Var}\left(\frac{N-1}{N} S^2\right)$$

$$= \left(\frac{N-1}{N}\right)^2 \operatorname{Var}(S^2)$$

$$= \left(\frac{N-1}{N}\right)^2 \sigma_{S^2}^2$$

$$= \left(1 - \frac{1}{N}\right)^2 \sigma_{S^2}^2$$

This is less than $\operatorname{Var}(S^2)$, since $1 - 1/N < 1$.

As is recalled, the relative efficiency of a statistic $\hat{\theta}_2$ to $\hat{\theta}_1$, where $\hat{\theta}_1$ is the efficient statistic, is given by

$$r_e = \frac{\operatorname{Var}(\hat{\theta}_1)}{\operatorname{Var}(\theta_2)}$$

In this case,

$$r_e = \frac{[(N-1)/N]^2 \sigma_{S^2}^2}{\sigma_{S^2}^2}$$

$$= \left(\frac{N-1}{N}\right)^2$$

For samples of size 5,

$$r_e = (\tfrac{4}{5})^2 = .64$$

Thus, S^2 is only 64 percent as efficient as $\hat{S}^2$. When $N = 25$,

$$r_e = (\tfrac{24}{25})^2 = .92$$

and by the time $N = 50$,

$$r_e = (\tfrac{49}{50})^2 = .96$$

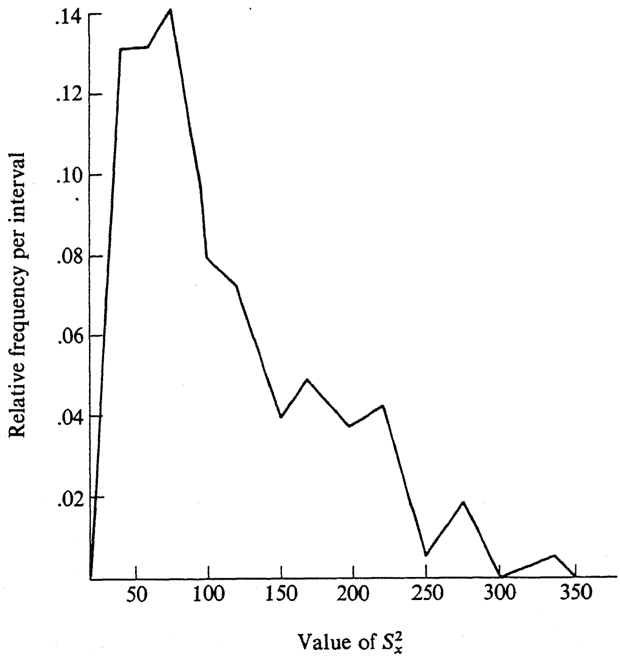
so that for large N, both estimates are nearly equal in efficiency.

Not only does the efficiency of S^2 approach $\hat{S}^2$ as N increases, but the bias in $\hat{S}^2$ reduces as N gets large. For example, when $N = 5$, the bias is $\beta = -\frac{1}{5}\sigma^2 = -.2\sigma^2$. When $N = 25$, the bias is $\beta = -\frac{1}{25}\sigma^2 = -.04\sigma^2$, and when $N = 50$, the bias is $\beta = -\frac{1}{50}\sigma^2 = .02\sigma^2$, so that for large N, $\hat{S}^2$ is almost unbiased as an estimator of σ^2.

It is true that for large samples it makes little difference what one uses as an estimator of σ^2, provided that the choice is restricted to S^2 and $\hat{S}^2$. However, the difference in these estimators is quite large for small samples and for that reason S^2 is recommended for all occasions. Later it will be seen that S^2 tends to simplify many of the formulas that appear in statistical inference problems or formulas, which is another reason why S^2 is preferred over $\hat{S}^2$.

The Monte Carlo generated sampling distribution of S^2 selected from the normal population $N(50,10^2)$ for samples of size 5 is shown in Figure 9-8. This distribution

Figure 9-8. Sampling distribution of S_X^2 for samples of size 5 from an approximately normal population with $\sigma_X^2 = 100$.



is not symmetrical. In fact, it is quite skewed and obviously not normal. As a first impression, one might suspect that because $E(S^2) = \sigma^2$, the $E(S)$ should be equal to σ. While this is intuitively plausible, it is false. For the 320 samples selected by the students in the beginning course in statistics,

$$\bar{S} = 8.76$$

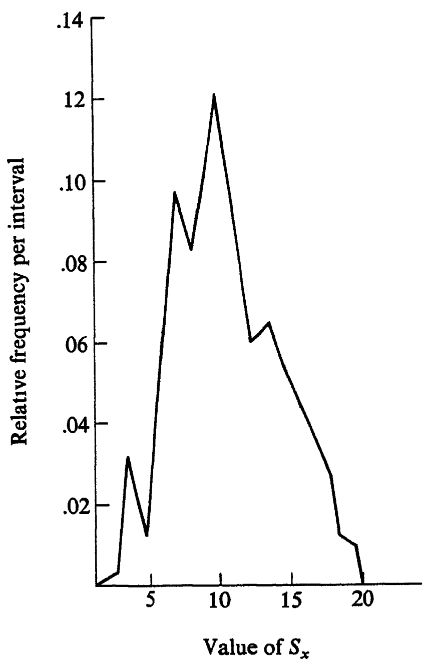
If the procedure were unbiased, then it would be true that $E(S) = 10$. The sample standard deviation is biased as an estimator of σ. It tends to give estimates of σ that are low, on the average. However, the amount of bias in S is much less than the bias in $\hat{S}$. For the 320 samples, the average value of $\hat{S}$ is given by $\bar{\hat{S}} = 6.70$, which is very far from the parameter value of 10. This is another reason why S is preferred over $\hat{S}$ as an estimator of σ.

The sampling distribution of S, for samples of size 5, selected from $N(50,10^2)$, is shown in Figure 9-9. This distribution is not symmetrical. However, as N increases, the sampling distribution of S tends to become normal, and for a good approximation,

$$E(S) \sim \sigma$$

$$\mathrm{Var}(S) \sim \frac{\sigma^2}{2N}$$

Figure 9-9. Sampling distribution of S_X for samples of size 5 from a population with $\sigma_X = 10$.



Even though the sampling distribution of S tends to normality, this property is not extensively used in the statistical analysis of data. Instead, a distribution closely allied to the sampling distribution of S^2, called the chi-square distribution, is employed.

9-10 THE SAMPLING DISTRIBUTION OF $\hat{p}$, THE ESTIMATOR OF THE BINOMIAL DISTRIBUTION PARAMETER

Consider the variable

$$T = X_1 + X_2 + \cdots + X_N$$

where X_i is Bernoulli with $P(X_i = 1) = p$ and $P(X_i = 0) = q$. As was shown in Theorems 6-2 and 6-5,

$$E(T) = Np$$

and

$$\operatorname{Var}(T) = Npq$$

Note that

$$\hat{p} = \frac{T}{N} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

has the form of a sample average. Therefore, what was previously stated about $\bar{X}$ must also be true for T/N. Since T is a weighted average of 1's and 0's, $\hat{p}$ equals the proportion of 1's in the sample.

Theorem 9-6

The sample proportion is an unbiased estimator of the population proportion, that is, $E(\hat{p}) = p$.

Proof. By definition and Theorem 7-3,

$$E(\hat{p}) = E\left(\frac{T}{N}\right) = \frac{1}{N}E(T) = \frac{1}{N}Np = p$$

This completes the proof.

Theorem 9-7

The variance of the sampling distribution of $\hat{p}$ is given by

$$\operatorname{Var}(\hat{p}) = \frac{1}{N}pq$$

Proof.

$$\operatorname{Var}(\hat{p}) = \operatorname{Var}\left(\frac{T}{N}\right) = \frac{1}{N^2}\operatorname{Var}(T)$$

$$= \frac{1}{N^2}Npq = \frac{pq}{N}$$

This completes the proof.

Since p is an *average* of 1's and 0's, it must behave according to the central limit theorem and indeed it does. As N becomes large, the sampling distribution of p approaches the normal distribution with parameters

$$E(\hat{p}) = p \qquad \text{and} \qquad \operatorname{Var}(\hat{p}) = \frac{pq}{N}$$

As an example of the use of this important result, consider a coin that is tossed 100 times and let $\hat{p}$ be the proportion of heads in the sample. Reasonable limits for $\hat{p}$ can be determined in a manner directly analogous to that used for means. Almost all of the probability distribution of $\hat{p}$ is concentrated between

$$\hat{p}: \{E(\hat{p}) - 3\sigma_{\hat{p}} < \hat{p} < E(\hat{p}) + 3\sigma_{\hat{p}}\}$$

If the coin is fair, $p = \frac{1}{2}$ and $E(\hat{p}) = p = \frac{1}{2}$ and $\operatorname{Var}(\hat{p}) = pq/N = \frac{1}{100}(\frac{1}{2})(\frac{1}{2}) = \frac{1}{400}$ so that

$$\sigma_{\hat{p}} = \tfrac{1}{20} = .05$$

and the expected range of $\hat{p}$ is

$$\hat{p}: \{.50 - 3(.05) < \hat{p} < .50 + 3(.05)\}$$

or

$$\hat{p}: \{.35 < \hat{p} < .65\}$$

If the coin were tossed and if $T = 45$, then $\hat{p} = .45$ is compatible with the hypothesis that the coin is fair. However, if the number of heads equalled 72, then $\hat{p} = .72$ is not compatible with the hypothesis that the coin is fair. In this case it is suspect.

Finally, since $\hat{p}$ is a simple average, it must be an efficient estimator of p. This last result is also of theoretical and applied interest. In discussing the relative-frequency interpretation of probability, it was said that if $\hat{p}$ came close to a number p, that number could be treated as though it were the probability of the event being studied. The question that immediately comes to mind is that of determining the value that N must take for the estimated value to be close to the unknown p. As an example, suppose that one wished to determine the probability of the event A so that the estimate would not deviate from the true value by more than .001. For large N, $\hat{p}$ is normal with $E(\hat{p}) = p$ and $\operatorname{Var}(\hat{p}) = pq/N$ so that the statement

$$\hat{p}: \{p - .001 < \hat{p} < p + .001\}$$

is equivalent to

$$p: \left\{ p - 3\sqrt{\frac{pq}{N}} < \hat{p} < p + 3\sqrt{\frac{pq}{N}} \right\}$$

so that

$$3\sqrt{\frac{pq}{N}} = .001$$

It is not difficult to show that pq/N is maximized when $p = \frac{1}{2}$. Thus, for this maximum value,

$$3\sqrt{\frac{(\frac{1}{2})(\frac{1}{2})}{N}} = \frac{1}{1000}$$

Solving this equation for N, we find that the number of observations required for this degree of precision is 2,250,000. For a less precise estimate, $p \pm .01$, the required sample is 22,500. For $p \pm .05$, the required sample size is 400. In any case, it is clear that exceptionally large samples are required to determine the probability of a dichotomous characteristic with a rather high degree of precision. In most studies, N is considerably less. However, as will be seen later, one can get by with much smaller samples provided one is willing to tolerate larger risks of being wrong.

9-11 THE VARIANCE OF THE SAMPLING DISTRIBUTION OF $\hat{p}$, THE ESTIMATOR OF A HYPERGEOMETRIC DISTRIBUTION PARAMETER

As was noted in Section 6-12, the variance of X for samples of size n from a hypergeometric distribution is given by

$$\mathrm{Var}(X) = npq\left(\frac{N-n}{N-1}\right)$$

From this, it follows that the $\mathrm{Var}(\hat{p})$ for a hypergeometric variable is given by

$$\mathrm{Var}(\hat{p}) = \mathrm{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\mathrm{Var}(X) = \frac{1}{n^2}npq\left(\frac{N-n}{N-1}\right) = \frac{pq}{n}\left(\frac{N-n}{N-1}\right)$$

Thus, one uses this formula for the sampling distribution of $\hat{p}$ if one is considering sampling with replacement from a finite universe. When $n/N < 10$ percent, one can use the binomial variance equation as a good approximation.

9-12 THE VARIANCE OF THE SAMPLING DISTRIBUTION OF $\bar{X}$ FOR FINITE UNIVERSES

The variance equation for $\hat{p}$ of the hypergeometric distribution is a special case of the following theorem.

Theorem 9-8

The variance of the sampling distribution of $\bar{X}$ for samples of size n selected from finite universes of size N is given by

$$\text{Var}(\bar{X}) = \frac{\sigma_X^2}{n}\frac{N-n}{N-1}$$

While proof of this result is not difficult, the basic theoretical results for the proof have not been presented in the earlier sections of this book. If the sampling fraction $n/N < 10$ percent, then one can ignore the result and use $\text{Var}(\bar{X}) = \sigma^2/n$, otherwise, the finite population correction should not be ignored.

9-13 SUMMARY

Since the parameters of a probability distribution are generally unknown, samples are selected and estimates are determined from the sample values. In general, there are unlimited numbers of formulas that can be generated to produce statistical estimates of parameters. Naturally, some estimators are better than others. In this chapter, two criteria were presented on which different estimators could be evaluated. These criteria relate to accuracy and precision over repeated experimentation or observation. If an estimation procedure gives rise to an accurate estimator, the estimator is said to be unbiased; if an estimation procedure gives rise to a precise estimator, the estimator is said to be efficient. While properties of estimators can be studied by means of an elegant mathematical theory, they were examined in this chapter via statistical sampling distributions generated by Monte Carlo procedures.

Over a sample space $\bar{X}$ and $\hat{M}$, the sample mean and median, vary so that each has a probability distribution. These probability distributions are called the sampling distributions of $\bar{X}$ and $\hat{M}$. If the original universe is normal in form with parameters μ and σ^2, then the sampling distributions of $\bar{X}$ and $\hat{M}$ are also normal in form with parameters given by $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \sigma^2/N$, and $\mu_{\hat{M}} = \mu$ and $\sigma_{\hat{M}}^2 = 1.56\sigma^2/N$. Since $E(\bar{X}) = \mu$ and $E(\hat{M}) = \mu$, the estimation procedures are unbiased. Since $\text{Var}(\bar{X}) < \text{Var}(\hat{M})$, it is said that $\bar{X}$ is the efficient estimator when compared to $\hat{M}$ as an estimator of μ. While it has not been shown, the sampling distribution of $\bar{X}$ is approximately normal even when X is not normal because of the influence of the central limit theorem.

For a general parameter θ that admits an estimator $\hat{\theta}$, if the sampling distribution of $\hat{\theta}$ has θ as its expected value, then $\hat{\theta}$ is said to be an unbiased estimator of θ. Since $E(S^2) = \sigma^2$, S^2 is an unbiased estimator of σ^2. In like manner, $\hat{p}$ is an unbiased estimator of p since $E(\hat{p}) = p$. While S^2 is an unbiased estimator of σ^2, S is a biased estimator of σ since $E(S) = \sigma + \beta$.

For a general parameter θ, a statistic, $\hat{\theta}_1$, is said to be a more efficient estimator of θ than another statistic, $\hat{\theta}_2$, if the variances of their sampling distributions are such that the $\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$. In this sense, $\bar{X}$ is the efficient estimator when

compared to $\hat{M}$. As a rough rule of thumb, a sample mean based on 64 observations is as efficient as a sample median based on 100 observations.

In general, one need not be too concerned over the selection of efficient estimators since most of the estimators employed in standard statistical procedures are efficient or almost efficient. This is true of S^2, which is not efficient but which is quite close to the efficient estimator if the sample is large.

While sampling distributions have been used to study the statistical properties of estimators, their major use is in making inferences from limited knowledge. This important use of sampling distributions will be developed in the remaining chapters.

EXERCISES

9-1. Explain what is meant by an unbiased estimator and an efficient estimator.

9-2. Explain what is meant by the sampling distribution of a statistic.

9-3. What are the reasons for preferring S to $\hat{S}$ as an estimator of σ, knowing that both estimators are biased?

***9-4.** Consider the experiment of Section 1-2 in which a psychologist wished to estimate the average IQ of children in a particular school district. For his study he planned on using $N = 50$ students If the population of students is like the general population, then it follows that X is $N(100, 16^2)$ Determine the 95 percent range of $\bar{X}$ for his sample average In his experiment it so happened that $\bar{X} = 89.3$. What does this suggest about his hypothesis?

***9-5.** Consider the experiment of Section 1-2 in the manner planned by the psychologist. Let the criterion variable be Y: {number of students with IQ below normal}. If the population of students is like the general population it follows that $P(X < 100) = \frac{1}{2}$ so that Y is $B(50,\frac{1}{2})$. Determine the 95 percent range for Y. In the experiment, $Y = 37$ What does this suggest about his hypothesis?

***9-6.** In Exercises 9-4 and 9-5, the psychologist felt that the performance would be low so that $\bar{X} < 100$ and $Y > 25$. This suggests that he should be concerned with the lower end of the $\bar{X}$ distribution and the upper end of the Y distribution. Determine these 95 percent ranges and answer the questions concerning his hypothesis.

***9-7.** In terms of the sampling distribution of means.

$$P\left(\mu - 1.96\frac{\sigma}{\sqrt{N}} < \bar{X} < \mu + 1.96\frac{\sigma}{\sqrt{N}}\right) = .95$$

so that

$$P\left(-1.96\frac{\sigma}{\sqrt{N}} < \bar{X} - \mu < 1.96\frac{\sigma}{\sqrt{N}}\right) = .95$$

Using this relationship, determine the sample size for the study of Exercise 9-4 if the deviation between the sample mean and population mean is to be less than 5 IQ points. For this determination, let $\sigma = 16$.

***9-8.** Following the format of Exercise 9-7, determine the sample size for the study if the criterion variable is given by Y: {number of students with IQ below 100} and the deviation between the sample proportion and the population proportion is to be less than .10.

9-9. Why is the sample mean based on 80 observations a better measure of μ than a sample median based on 100 observations? How many observations are needed to make the sample median as good an estimator as a sample mean based on 80 observations?

9-10. When conducting a study and computing $\bar{X}$, there is only one mean. How does this mean relate to the sampling distribution of $\bar{X}$? Explain.

10 THE CHI-SQUARE AND *t* DISTRIBUTIONS

NEYMAN RECEIVES NATIONAL MEDAL OF SCIENCE

Jerzy Neyman, 1968 winner of the Samuel S. Wilks Memorial Medal of the American Statistical Association, has been awarded the National Medal of Science in the Mathematical Sciences. This is the Federal Government's highest award for distinguished achievement The citation accompanying the award stated to "Jerzy Neyman, Professor of Mathematics, University of California, Berkeley. For laying the foundations of modern statistics and devising tests and procedures that have become essential parts of the knowledge of every statistician"

By permission from *The American Statistician*, February 1969, the American Statistical Association, Washington, D.C.

10-1 INTRODUCTION TO THE CHI-SQUARE AND t DISTRIBUTIONS

While the normal distribution dominates elementary statistical theory, there are other distributions that are also of importance in the study of observational and experimental data Two of these distributions are presented in this chapter. These closely related distributions are called the chi-square and t distributions. The chi-square distribution was introduced by Karl Pearson in 1900, while the t distribution was introduced in 1908 by William Gosset, writing under the pen name Student. In some circles, this last-named distribution is referred to as Student's t distribution.

Both of these distributions are defined in terms of only one parameter. This parameter is usually denoted by the Greek letter ν (nu) and is named the degrees of freedom of the distributions. The meaning or rationale of this name will be explained in the following section.

10-2 DEGREES OF FREEDOM ASSOCIATED WITH A SAMPLE VARIANCE

Before attempting a definition of ν, recall Theorem 6-3, which states that the sum of the weighted deviations about the expected value of a discrete random variable is equal to 0. As might be expected, this same theorem is valid for continuous random variables and has an analogous counterpart in a set of sample values. Since a sample average is analogous to a population expected value, it might be supposed that the sum of the weighted deviations about a sample average is equal to 0. The truth of this generalization is shown in the following theorem for unweighted deviations.

Theorem 10-1

The sum of the deviations about a sample average is 0. Stated algebraically,

$$\sum_{i=1}^{N} d_i = \sum_{i=1}^{N} (X_i - \bar{X}) = 0$$

Proof. By definition, the sum of the deviations about $\bar{X}$ is given by

$$\begin{aligned}\sum_{i=1}^{N} d_i = \sum_{i=1}^{N} (X_i - \bar{X}) &= (X_1 - \bar{X}) + (X_2 - \bar{X}) + \cdots + (X_N - \bar{X}) \\ &= (X_1 + X_2 + \cdots + X_N) - (\bar{X} + \bar{X} + \cdots + \bar{X}) \\ &= \sum_{i=1}^{N} X_i - N\bar{X} \\ &= N\bar{X} - N\bar{X} \\ &= 0\end{aligned}$$

This completes the proof.

As an illustration of the use of this theorem, consider the sample X: {3,8,14,6,9} for which $\bar{X} = 8$. For this sample,

$$d_1 = (X_1 - \bar{X}) = 3 - 8 = -5$$

$$d_2 = (X_2 - \bar{X}) = 8 - 8 = 0$$

$$d_3 = (X_3 - \bar{X}) = 14 - 8 = 6$$

$$d_4 = (X_4 - \bar{X}) = 6 - 8 = -2$$

$$d_5 = (X_5 - \bar{X}) = 9 - 8 = 1$$

The sum of these deviations is given by

$$\sum_{i=1}^{5} d_i = (-5) + (0) + (6) + (-2) + (1) = 0$$

While it is true that the deviations add to 0, another important property concerning these deviations is worthy of note and this property is that the differences are not independent. As soon as four are known, the fifth is determined. For example, suppose that $d_1 = -5$, $d_2 = 0$, $d_3 = 6$, and $d_5 = 1$ are known. It follows from Theorem 10-1 that d_4 is also known, since $d_1 + d_2 + d_3 + d_4 + d_5 = 0$. Substituting the numerical values of the known deviations into this equation, we find that

$$-5 + 0 + 6 + d_4 + 1 = 0$$

so that

$$d_4 = 5 - 0 - 6 - 1 = -2$$

which it is. In a case like this, it is said that the five deviations possess 4 degrees of freedom. When any four are known, the last one is determined. If the sample has N elements, the number of independent deviations is denoted by $\nu = (N - 1)$.

While it is not explicitly stated, this concept of degrees of freedom has been utilized earlier when the definition of a sample variance was introduced. As is recalled, the sample variance was defined in terms of $(N - 1)$ instead of N. By definition, the sample variance is given by

$$S_X^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N - 1}$$

The number of independent deviations is equal to $\nu = (N - 1)$, which is also the divisor for estimating the variance. Under this discussion, S_X^2 is thus said to be an estimate of σ_X^2 based upon $\nu = (N - 1)$ degrees of freedom.

Suppose that the population from which the sample X: {3,8,14,6,9} was taken had an expected value equal to μ. Then the sample deviations from μ are:

$$d_1^* = (3 - \mu)$$
$$d_2^* = (8 - \mu)$$
$$d_3^* = (14 - \mu)$$
$$d_4^* = (6 - \mu)$$
$$d_5^* = (9 - \mu)$$

In this case knowledge concerning four of the deviations reveals absolutely no information about the fifth deviation, since

$$d_1^* + d_2^* + d_3^* + d_4^* + d_5^*$$

need not equal 0. Thus, an estimate of σ^2 based upon the d_i^* would be an estimate based upon $\nu = N$ degrees of freedom. A formula for this estimate would be given by

$$S^{*2} = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

which is similar to the formula of the variance of a discrete probability distribution where each value has the same probability of occurrence.

While random variables of the form $T = X_1 + X_2 + \cdots + X_N$ and $T' = (X_1 - \bar{X}) + (X_2 - \bar{X}) + \cdots + (X_N - \bar{X})$ assume the dominant role of statistical theory and tend to approximate a normal distribution, random variables of the form $U = X_1^2 + X_2^2 + \cdots + X_N^2$ and $U' = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_N - \bar{X})^2$ also play an important role in statistical theory. Variables of this nature tend to approximate a distribution called the chi-square distribution. That they are not normal in form should not come as a surprise since it was noted in Chapter 9 that for $N = 5$, the sampling distribution of

$$S_X^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_N - \bar{X})^2}{N - 1} = \frac{U'}{N - 1}$$

was quite skewed and far from normal in form.

10-3 THE CHI-SQUARE DISTRIBUTION

While the chi-square distribution is a unique distribution, it is intimately associated with the $N(0,1)$ distribution. In fact, the chi-square distribution is a probability distribution generated from the $N(0,1)$ distribution in the manner stated in the following theorem.

Theorem 10-2

If a variable Z has a $N(0,1)$ probability distribution, then Z^2 has a chi-square distribution with 1 degree of freedom: $Z^2 = \chi_1^2$.

Unfortunately, the proof of this theorem is beyond the scope of this book and its truth must be accepted on faith. In any case, it should be noted that χ_1^2 can never assume a negative value since it represents the probability distribution of the square of a variable. Since Z has its greatest probability near 0, χ_1^2 values near 0 have greatest probability. The distribution has the form shown in Figure 10-1.

Two properties of the chi-square distribution with 1 degree of freedom are summarized in the following two theorems, which are stated without proofs.

Theorem 10-3

The expected value of the chi-square distribution with 1 degree of freedom is equal to 1:

$$E(\chi_1^2) = 1$$

Theorem 10-4

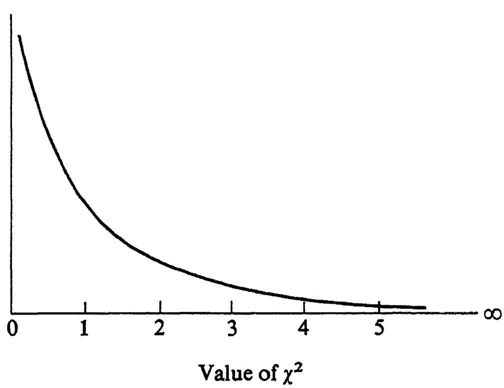
The variance of the chi-square distribution with 1 degree of freedom is equal to 2:

$$\mathrm{Var}(\chi_1^2) = 2$$

Percentiles of this distribution are very easy to compute, using tables of Z. In computing these percentiles, one notes that $(-Z)^2 = Z^2$. This means that all Z values below 0 become positive χ_1^2 values when squared. Thus

$$P(\chi_1^2 > a^2) = P(Z < -a) + P(Z > a)$$

Figure 10-1. The chi-square distribution with 1 degree of freedom



Thus, if $a = 1.96$ and $-a = -1.96$, then $a^2 = (1.96)^2 = 3.84$, and $(-a)^2 = (-1.96)^2 = 3.84$, so that

$$\begin{aligned} P(\chi_1^2 > 3.84) &= P(Z < -1.96) + P(Z > 1.96) \\ &= .025 + .025 \\ &= .05 \end{aligned}$$

As another example, if $a = 1.645$ and $-a = -1.645$, then $a^2 = (1.645)^2 = 2.71$ and $(-a)^2 = (-1.645)^2 = 2.71$, so that

$$\begin{aligned} P(\chi_1^2 > 2.71) &= P(Z < -1.645) + P(Z > 1.645) \\ &= .05 + .05 \\ &= .10 \end{aligned}$$

Using the chi-square with 1 degree of freedom, one can generate other probability distributions that play an important role in the study of variances. These results are stated as theorems. A proof of Theorem 10-5 is not given.

Theorem 10-5

If $Z_1, Z_2, \ldots, Z_N$ are independent, each $N(0,1)$, then $Z_1^2 + Z_2^2 + \cdots + Z_N^2$ has a chi-square distribution with $\nu = N$ degrees of freedom.

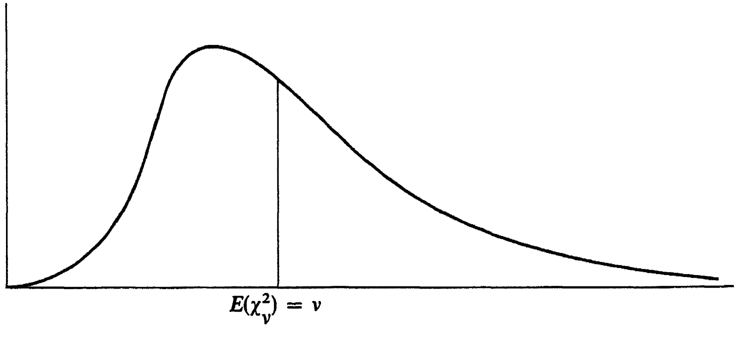
The general form of this distribution is as shown in Figure 10-2.

Theorem 10-6

When N is small the distribution of χ_N^2 is extremely skewed, but as N increases the distribution approaches normal form, with

$$E(\chi_N^2) = N \qquad \text{and} \qquad \operatorname{Var}(\chi_N^2) = 2N$$

Figure 10-2 General form of the chi-square distribution with ν degrees of freedom



Proof. The approach to normality follows from the central limit theorem since $U = X_1^2 + X_2^2 + \cdots + X_N^2$ is of the basic form $T = Y_1 + Y_2 + \quad + Y_N$. By definition,

$$E(\chi_N^2) = E(Z_1^2 + Z_2^2 + \cdot \quad \cdot + Z_N^2)$$

By Theorem 6-1,

$$E(\chi_N^2) = E(Z_1^2) + E(Z_2^2) + \quad \cdot \quad + E(Z_N^2)$$

By Theorem 10-3, $E(\chi_1^2) = 1$, so that

$$E(\chi_N^2) = 1 + 1 + \cdot \quad \cdot + 1 = N$$

By definition,

$$\begin{aligned} \mathrm{Var}(\chi_N^2) &= \mathrm{Var}(Z_1^2 + Z_2^2 + \cdots + Z_N^2) \\ &= \mathrm{Var}(Z_1^2) + \mathrm{Var}(Z_2^2) + \quad \cdot\cdot + \mathrm{Var}(Z_N^2) \end{aligned}$$

By Theorem 10-4,

$$\mathrm{Var}(\chi_1^2) = 2$$

so that

$$\mathrm{Var}(X_N^2) = 2 + 2 + \cdots + 2 = 2N$$

This completes the proof.

Note that the expected value and variance are defined in terms of N, the degrees of freedom of the distribution. This number, as stated earlier, is the parameter of the chi-square distribution.

Theorem 10-5 illustrates an important property of the chi-square distribution that gives rise to a set of procedures embodied in the analysis of variance. This property is termed the additive property of χ^2. As an illustration of this property, let

$$\chi_8^2 = Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 + Z_5^2 + Z_6^2 + Z_7^2 + Z_8^2$$

By definition,

$$Z_1^2 + Z_2^2 + Z_3^2 = \chi_3^2$$

and

$$Z_4^2 + Z_5^2 + Z_6^2 + Z_7^2 + Z_8^2 = \chi_5^2$$

Thus,

$$\chi_8^2 = \chi_3^2 + \chi_5^2$$

This illustrates the following theorem, which is stated without proof.

Theorem 10-7

If $U_1 = \chi^2_{\nu_1}$ and $U_2 = \chi^2_{\nu_2}$ and if U_1 and U_2 are statistically independent, then

$$U_3 = U_1 + U_2 = \chi^2_{\nu_1} + \chi^2_{\nu_2}$$

has a chi-square distribution with parameter

$$\nu_3 = \nu_1 + \nu_2$$

This last result will now be used to conclude the study of the sampling distribution of S^2 begun in Chapter 9.

10-4 THE SAMPLING DISTRIBUTION OF S^2 (CONTINUATION)

Consider a random sample of size N drawn from a normal population with $E(X) = \mu$ and $\operatorname{Var}(X) = \sigma^2$. Let the sample be denoted as $(X_1, X_2, \ldots, X_N)$. From Theorem 7-7,

$$Z_1 = \frac{X_1 - \mu}{\sigma} \quad \text{has a } N(0,1) \text{ distribution}$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} \quad \text{has a } N(0,1) \text{ distribution}$$

$$\vdots$$

$$Z_N = \frac{X_N - \mu}{\sigma} \quad \text{has a } N(0,1) \text{ distribution}$$

According to Theorem 10-5,

$$Z_1^2 + Z_2^2 + \cdots + Z_N^2 = \left(\frac{X_1 - \mu}{\sigma}\right)^2 + \left(\frac{X_2 - \mu}{\sigma}\right)^2 + \cdots + \left(\frac{X_N - \mu}{\sigma}\right)^2$$

has a chi-square distribution with N degrees of freedom. Thus

$$\chi^2_N = \sum_{i=1}^{N} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{N} (X_i - \mu)^2$$

Using the additive property of χ^2, we can partition

$$\chi^2_N = \frac{1}{\sigma^2} \sum_{i=1}^{N} (X_i - \mu)^2$$

into two components with $\nu_1 = (N - 1)$ and $\nu_2 = 1$ degrees of freedom. Furthermore, these components can be related to S_X^2 and $\bar{X}$, the sample estimates of σ_X^2 and μ. Stated as a theorem, we have the following:

Theorem 10-8

$$\chi^2_N = \sum_{i=1}^{N} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{(N-1)\,S_X^2}{\sigma^2} + \frac{N(\bar{X} - \mu)^2}{\sigma^2} = \chi^2_{N-1} + \chi^2_1$$

Proof. Using the binomial expansion, note that

$$\sum_{i=1}^{N} (X_i - \mu)^2 = \sum_{i=1}^{N} [(X_i - \bar{X}) + (\bar{X} - \mu)]^2$$

$$= \sum_{i=1}^{N} [(X_i - \bar{X})^2 + 2(X_i - \bar{X})(\bar{X} - \mu) + (\bar{X} - \mu)^2]$$

$$= \sum_{i=1}^{N} (X_i - \bar{X})^2 + 2(\bar{X} - \mu) \sum_{i=1}^{N} (X_i - \bar{X}) + N(\bar{X} - \mu)^2$$

$$= (N = 1) S^2 + 2(\bar{X} - \mu)(0) + N(\bar{X} - \mu)^2$$

Thus

$$\chi_N^2 = \frac{(N-1) S^2}{\sigma^2} + \frac{N}{\sigma^2}(\bar{X} - \mu)^2$$

Consider the second term on the right-hand side of this equation. Employing some simple algebra, we see that

$$\frac{N(\bar{X} - \mu)^2}{\sigma^2} = \frac{(\bar{X} - \mu)^2}{\sigma^2/N} = \left[\frac{(\bar{X} - \mu)}{\sigma/\sqrt{N}}\right]^2 = \left[\frac{(\bar{X} - \mu)}{\sigma_{\bar{X}}}\right]^2$$

$$= Z^2 = \chi_1^2$$

Thus

$$\chi_N^2 = \frac{(N-1) S^2}{\sigma^2} + \chi_1^2$$

By subtraction and the use of Theorem 10-7, we find that $(N - 1)S^2/\sigma^2$ has a chi-square distribution with $(N - 1)$ degrees of freedom Thus

$$\chi_{N-1}^2 = \frac{(N-1) S^2}{\sigma^2}$$

and

$$\chi_N^2 = \chi_{N-1}^2 + \chi_1^2$$

This completes the proof.

Two frequently used equations have been generated by the algebra of the proof of this theorem. These important equations are:

$$\chi_1^2 = \frac{(\bar{X} - \mu)^2}{\sigma_{\bar{X}}^2} = \frac{(\bar{X} - \mu)^2}{\sigma^2/N}$$

and

$$\chi_{N-1}^2 = \frac{(N-1) S^2}{\sigma^2}$$

From the last equation it is seen that

$$S^2 = \frac{\sigma^2}{N-1}\chi_{N-1}^2$$

This indicates that the sampling distribution of S^2 is just a constant multiple of the χ^2 distribution with $\nu = N - 1$ degrees of freedom. This accounts for the extreme skewness in the sampling distribution of the sample variance shown in Figure 9-9.

As an example of the use of this sampling distribution, consider an experiment in which 25 college sophomores are to participate and suppose it is known from past experience that $\sigma^2 = 40$. One might wonder as to what are reasonable limits to place on the sample variance. If reasonable limits are defined to include the central 99 percent of the distribution, one can determine a 99 percent range for S^2 as follows. Since $N = 25$, $\nu = N - 1 = 24$. In Table A-7, which shows the χ^2 distribution, for $N = 24$ one reads that $\chi^2_{24}(.005) = 9.89$ and $\chi^2_{24}(.995) = 45.56$. By definition,

$$P[\chi^2_\nu(.005) < \chi^2 < \chi^2_\nu(.995)] = .99$$

would give reasonable limits of expectation. Thus, the smallest variance expected is given by

$$S^2_{\text{Smallest}} = S^2(.005) = \frac{\sigma^2\,\chi^2_\nu(.005)}{N-1} = \frac{40(9.89)}{24} = 16.48$$

The largest variance expected is

$$S^2_{\text{Largest}} = S^2(.995) = \frac{\sigma^2\,\chi^2_\nu(.995)}{N-1} = \frac{40(45.56)}{24} = 75.93$$

Thus, the expected 99 percent range for S^2 is given by

$$16.48 < S^2 < 75.93$$

If the experiment is now performed and it is found that $S^2 = 62.14$, then one would have no reason to believe that σ^2 was not equal to 40.

Suppose the experiment were now repeated but instead of 25 subjects, 400 were used. Since $\nu = 120$ is the last value shown in the χ^2 table, some other procedure must be used to determine $\chi^2_{399}(.005)$ and $\chi^2_{399}(.995)$. A very convenient way for estimating these values is given by the following formula:

$$\chi^2_\nu(Q) = \tfrac{1}{2}(Z_Q + \sqrt{2\nu - 1})^2$$

where Z_Q is the Qth percentile of the $N(0,1)$ distribution. For this example, $\nu = 399$, and

$$Z(.005) = -2.576 \qquad \text{and} \qquad Z(.995) = 2.576$$

so that

$$\begin{aligned}\chi^2_{399}(.005) &= \tfrac{1}{2}(-2.576 + \sqrt{2(399) - 1})^2\\ &= \tfrac{1}{2}(25.66)^2\\ &= 329.22\end{aligned}$$

and

$$\begin{aligned}\chi^2_{399}(.995) &= \tfrac{1}{2}(2.576 + \sqrt{2(399) - 1})^2 \\ &= \tfrac{1}{2}(30.81)^2 \\ &= 474.63\end{aligned}$$

Thus, the smallest sample variance to be expected is

$$S^2_{\text{Smallest}} = S^2(.005) = \frac{\sigma^2 \chi^2_\nu(.005)}{N-1} = \frac{40(329.22)}{399} = 33.00$$

and the largest sample variance to be expected is

$$S^2_{\text{Largest}} = S^2(.995) = \frac{\sigma^2 \chi^2_\nu(.995)}{N-1} = \frac{40(474.63)}{399} = 47.58$$

Thus, the expected 99 percent range for S^2 is given by

$$33.00 < S^2 < 47.58$$

If the sample variance based on 400 observations had been 62.14, then one would think that the variability of the experiment was larger than expectation, or that σ^2 was considerably larger than 40.

10-5 THE t DISTRIBUTION

In the previous section it was seen that

$$\chi^2_N = \frac{1}{\sigma^2} \sum_{i=1}^{N} (X_i - \mu)^2$$

could be partitioned into two components such that

$$\frac{\sum(X_i - \mu)^2}{\sigma^2} = \frac{(N-1) S^2}{\sigma^2} + \frac{(\bar{X} - \mu)^2}{\sigma^2/N}$$

and

$$\chi^2_N = \chi^2_{N-1} + \chi^2_1$$

indicating that

$$\frac{(N-1) S^2}{\sigma^2} \quad \text{is distributed as } \chi^2_{N-1}$$

and

$$\frac{(\bar{X} - \mu)^2}{\sigma^2/N} \quad \text{is distributed as } \chi^2_1$$

Even though these two distributions are statistically independent, they can be related to one another by a ratio involving $\bar{X}$ and S, the sample mean and sample standard deviation. This relationship was developed by Gosset through the simple ratio consisting of the square roots of the two chi-square components. In particular, if the following ratio is examined it is seen that

$$t = \frac{\sqrt{\chi_1^2/1}}{\sqrt{\chi_{N-1}^2/(N-1)}}$$

$$= \frac{\sqrt{\dfrac{\dfrac{(\bar{X}-\mu)^2}{\sigma^2/N}}{1}}}{\sqrt{\dfrac{(N-1)\,S^2/\sigma^2}{N-1}}}$$

$$= \frac{\dfrac{(\bar{X}-\mu)}{\sigma/\sqrt{N}}}{S/\sigma}$$

$$= \frac{\bar{X}-\mu}{S/\sqrt{N}}$$

This equation looks almost like

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$$

The only difference between this new variable t and the variable Z is the substitution of S for σ. Since σ is generally unknown, the utility of such a variable is obvious. This new variable is called a t variable, or a Student t variable in honor of Gosset. This variable has a distribution that is very similar to the normal. It is symmetrical, its expected value is given by $E(t) = 0$, and its variance is given by

$$\mathrm{Var}(t) = \frac{\nu}{\nu - 2}$$

where ν = number of degrees of freedom for estimating σ. When $N = 5$, $\mathrm{Var}(t) = 4/(4-2) = 2$, so that the spread in t is larger than the spread in Z. When $N = 30$, $\mathrm{Var}(t) = \frac{29}{27} = 1.0741$, so that the $\sigma_t = 1.0364$ and the distribution of t is very close to normal; one could therefore use the normal curve to approximate the t curve when $N \geqslant 30$.

As an example of the use of this statistic, suppose a random sample of 25 students is given a reading test for which the published norms state that $\mu = 80$ and $\sigma = 15$. Suppose also that the norms are based upon middle-class white students

enrolled in suburban schools. If the 25 students are not members of this population, application of the published norms to their scores is suspect. Suppose the students are enrolled at a school in a high socioeconomic neighborhood where test performance is quite uniform, so that $\sigma = 15$ is actually too large. If one uses the sample standard deviation as an estimate of σ, one might like to know what are reasonable limits to expect for average performances in similar samples of students. If the sample standard deviation is given by $S = 8.2$, then $\nu = N - 1 = 25 - 1 = 24$. Table A-8 shows that $t_{24}(.005) = -2.797$ and $t_{24}(.995) = 2.797$. With these values it is seen that $\bar{X}_{\text{Smallest}}$ and $\bar{X}_{\text{Largest}}$ are defined by the following two equations:

$$t_S = t_\nu(.005) = \frac{\bar{X}_S - \mu}{S/\sqrt{N}} \qquad t_L = t_\nu(.995) = \frac{\bar{X}_L - \mu}{S/\sqrt{N}}$$

$$-2.797 = \frac{\bar{X}_S - 80}{8.2/\sqrt{25}} \qquad 2.797 = \frac{\bar{X}_L - 80}{8.2/\sqrt{25}}$$

or

$$\bar{X}_S = 80 - 2.797\frac{8.2}{\sqrt{25}} \qquad \bar{X}_L = 80 + 2.797\frac{8.2}{\sqrt{25}}$$

$$= 80 - 4.49 \qquad = 80 + 4.49$$

$$= 75.51 \qquad = 84.49$$

Thus, a reasonable range for $\bar{X}$ in other similar classes is given by

$$75.5 < \bar{X} < 84.5$$

Further uses of this distribution are presented in the remaining chapters.

10-6 SAMPLING DISTRIBUTIONS AND PARAMETER INFERENCES

In Chapter 9, it was shown how the sampling distribution of the means, known to be approximately normal in form, could be used to make inferential statements about population expected values. This was done by finding reasonable ranges for the sample average and then checking to see whether the observed sample average fitted in the expected range. If it did not, then it was reasonable to suspect either the hypothesis or the assumptions that were made in computing the expected range. Additionally, it was shown how the sampling distribution of $\hat{p}$, also known to be approximately normal in form, could be used to make inferences about p, and now it has been shown how the χ^2 distribution can be used to accomplish the same goal for σ^2. In addition, the use of the t distribution with respect to μ when σ^2 is unknown was also shown. Unfortunately, it was necessary to assume that the $E(X)$ was known before inferences about the expected value could be made. Clearly, this is an unreasonable assumption, for if it were known there would be no need to estimate it or make inferences about it. The questions that were answered, while being interesting, are not very realistic and are really quite silly and unimportant.

Instead of asking what is a reasonable range for $\bar{X}$, the sample average, it makes more sense to ask what is a reasonable range for μ, the population expected value. As the following example will show, this question is more interesting for several reasons.

Suppose a new Ph.D. conducts a study on reading using fourth-grade school children, and as a measuring instrument he uses a standard achievement test. Suppose that the sample size of the study is equal to 100 and that the average score of the students is 80.1. Suppose also that exactly the same study is performed in another school district in a school with similar students and facilities by an older, famous educator who decides to use a sample size of 400, and suppose for his subjects the average score is given by 70.1. The immediate question that comes to mind is "Who is right—the older famous researcher or the new Ph.D.?" Those uninitiated in statistical methodology might tend to believe the established researcher because he is famous and because he used four times as many students as the new Ph.D. Such reasoning is not recommended because both could be correct, since sample averages are not generally equal to the population expected value. Because of this, one would like to be able to account for the random variation that exists between two numbers that presumably are measures of the same unknown parameter. If this could be done, one might find that there is no *real* statistical difference between the two averages obtained in the two experiments. The problem posed here is a very common one that will be given considerable attention in the next chaper.

10-7 SUMMARY

In this chapter, two important statistical probability distributions were introduced, the chi-square and t-distribution. Both distributions are defined in terms of one parameter, called the degree of freedom and denoted by ν. This parameter appears in the denominator of the formula for a sample variance.

Both the χ^2_ν and t_ν distributions are generated from the normal distribution. In particular, if $\{X_1, X_2, \ldots, X_N\}$ are independent, each $N(\mu, \sigma^2)$, then

$$\chi^2_N = \sum_{i=1}^{N} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{N} (X_i - \mu)^2$$

If μ is replaced by $\bar{X}$, then

$$\chi^2_{N-1} = \sum_{i=1}^{N} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{N} (X_i - \bar{X})^2 = \frac{(N-1)\,S^2}{\sigma^2}$$

One of the classical equations of statistical theory is given by

$$\sum_{i=1}^{N} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 = \frac{(N-1)\,S^2}{\sigma^2} + \frac{N(\bar{X} - \mu)^2}{\sigma^2}$$

from which it is seen that

$$\chi^2_N = \chi^2_{N-1} + \chi^2_1$$

The importance of this equation is connected to the way it relates S^2 and $\bar{X}$ to the χ^2_{N-1} and χ^2_1 distributions. From this last equation, the probability distribution of t_{N-1} is denoted as

$$t_{N-1} = \sqrt{\frac{\chi^2_1/1}{\chi^2_{N-1}/(N-1)}} = \frac{\bar{X}-\mu}{S/\sqrt{N}}$$

Both the χ^2_ν and t_ν play a major role in the analysis of experimental and observational data. Unfortunately, their importance was not sufficiently emphasized in this chapter. However, their application to statistical inference procedures will be demonstrated repeatedly in the chapters that follow.

One application of the χ^2 distribution to statistical questions related to sample variances was demonstrated. In particular, it was shown that if X is $N(\mu,\sigma^2)$, then the sampling distribution of S^2 is equal to χ^2 multiplied by $\sigma^2/(N-1)$. From this property, a $(1-\alpha)$ percent central range for S^2 was shown to be given by

$$\frac{\sigma^2}{N-1}\chi^2_\nu\left(\frac{\alpha}{2}\right) < S^2 < \frac{\sigma^2}{N-1}\chi^2_\nu\left(1-\frac{\alpha}{2}\right)$$

In a like manner, it was seen that if X is $N(\mu,\sigma^2)$ and if σ^2 is unknown, one can determine a $(1-\alpha)$ percent central range for $\bar{X}$ by

$$\mu - t_\nu\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{N}} < \bar{X} < \mu + t_\nu\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{N}}$$

Other more important statistical uses of χ^2_ν and t_ν will be demonstrated in the remaining chapters Many of these techniques owe their existence to Professor Emeritus Jerzy Neyman of the University of California, Berkeley Much of the material of Chapters 11 and 12 is based upon his contributions to statistical theory. It was his genius that gave rise to the important statistical concepts of confidence intervals, power, and modern statistical hypothesis testing. These three concepts play the dominating role in the remaining chapters of this text.

Modern statistical theory has had many able men contribute to its growth, but five men stand out as its greatest contributors. These men are Frederick Gauss (1777–1855), Karl Pearson (1857–1936), William Gosset (1876–1937), Sir Ronald Fisher (1890–1962), and Jerzy Neyman (1894–   ). Gauss contributed to the study of the normal distribution and method of least squares; Pearson introduced the chi-square distribution and the correlation coefficient, a measure of association between two variables; Gosset introduced the t distribution and is often called the father of small-sample theory, Fisher contributed to the theory of interval estimation and to statistical procedures included in the broad topic called analysis of variance; and Neyman's greatest contributions are in statistical test theory, confidence intervals, and statistical power.

EXERCISES

10-1. In Table A-7 find

(*a*) $\chi^2_3(.005)$
(*b*) $\chi^2_7(.01)$
(*c*) $\chi^2_{10}(.025)$
(*d*) $\chi^2_{13}(.05)$
(*e*) $\chi^2_{15}(.90)$
(*f*) $\chi^2_{20}(.95)$
(*g*) $\chi^2_{40}(.975)$
(*h*) $\chi^2_{60}(.99)$
(*i*) $\chi^2_{120}(.995)$

10-2. Determine

(*a*) $\chi^2_{180}(.005)$
(*b*) $\chi^2_{300}(.90)$
(*c*) $\chi^2_{440}(.95)$

10-3. In Table A-8 find

(*a*) $t_4(.10)$
(*b*) $t_9(.25)$
(*c*) $t_{18}(.50)$
(*d*) $t_{20}(.75)$
(*e*) $t_{23}(.95)$
(*f*) $t_{29}(.975)$
(*g*) $t_{45}(.99)$
(*h*) $t_{90}(.995)$

10-4. Determine the 25th and 75th percentile of the χ^2_1 distribution.

10-5. Define or explain what is meant by a chi-square variable. Do the same for a t variable.

10-6. Under what conditions can the χ^2_ν and t_ν be approximated by a $N(0, 1)$ variable?

10-7. If a variable is $N(\mu,36)$, find the central 95 percent range for S^2 if

(*a*) S^2 is based on 10 observations
(*b*) S^2 is based on 20 observations
(*c*) S^2 is based on 40 observations

What does this tell you about S^2 as an estimate of σ^2 as the sample size is increased?

10-8. If a variable is $N(25,\sigma^2)$, find the central 95 percent range for $\bar{X}$ if

(*a*) $\bar{X}$ is based on 10 observations and $S^2 = 36$
(*b*) $\bar{X}$ is based on 20 observations and $S^2 = 36$
(*c*) $\bar{X}$ is based on 40 observations and $S^2 = 36$

What does this tell you about $\bar{X}$ as an estimate of μ as the sample size increases?

10-9. List the similarities and differences between

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \quad \text{and} \quad t = \frac{\bar{X} - \mu}{S/\sqrt{N}}$$

10-10. The publishers of a frequently employed achievement test claim that the average score on the test is equal to 80 with a standard deviation of 20. Determine the 95 percent range for $\bar{X}$ and S^2 for samples of size 24. This test was given to 24 students for which the resulting statistics were $\bar{X} = 86.2$ and $S^2 = 528.3$. What do these statistics suggest to you about the students tested?

11 INTERVAL ESTIMATION

The careful statistician hedges his conclusions with a statement which isolates and measures the uncertainty that is inherent in any statistical analysis. He would never be caught, for example, concluding flatly from a sample that the average income of a certain large group of people was, say, $8,000. He would carefully append some modest reservation such as: "If I repeated my procedure many times, in only 5 percent of the cases would my estimate differ from the true average by more than $200." This scrupulous definition of the degree of sureness is the distinguishing mark of modern statistical analysis.

11-1 INTRODUCTION TO INTERVAL ESTIMATION (PRECISION OF POINT ESTIMATES)

In Chapter 8, point-estimation procedures were introduced for μ, σ^2, and p. The estimators generated, namely, $\bar{X}$, $\hat{M}$, S^2, and $\hat{p}$, are called point estimates since they produce a specific numerical value that is used as an estimate of the unknown parameters. In Chapter 9, it was seen that point estimators vary over repeated sampling. As a result, it would seem that this natural variability could be incorporated into the estimation procedures. One way to do this is to specify a set of numerical values that could serve as possible estimates of a population parameter. Such sets of numbers generate interval estimates. These interval estimates are called confidence intervals. To help motivate the need for confidence intervals or estimates for population parameters, consider the following problem.

Consider a college administrator who would like to determine whether the intellectual capacity of the students enrolled in the college is increasing over time. For this assessment, suppose that he decides to select a random sample of 25 college students and to use their average College Board test scores as an estimate of μ, the average score of the students at the college. Before the sample is taken, one may make reasonable guesses as to the largest and smallest sample mean that could occur under random sampling. Suppose that a similar study was made in 1960 and suppose that it was noted at the time of that study that $\mu = 600$ and $\sigma = 80$. If there has been no change in the distribution of College Board test scores of students who enroll at the college, then the sampling distribution of means generated from this population will be reasonably close to normal with

$$\mu_{\bar{X}} = E(\bar{X}) = E(X) = 600$$

and

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{N}} = \frac{80}{\sqrt{25}} = \frac{80}{5} = 16$$

If the administrator wants to be reasonably certain about the range of values possible for $\bar{X}$, he could set up limits based upon $E(\bar{X}) \pm 3\sigma_{\bar{X}}$. However, such an interval would be exceptionally wide and of little practical utility. If the administrator is somewhat liberal in regard to the precision of the estimation, then intervals based upon $E(\bar{X}) \pm 1.96\sigma_{\bar{X}}$ are quite reasonable. With this concession, it follows that

$$P[E(\bar{X}) - 1.96\sigma_{\bar{X}} < \bar{X} < E(\bar{X}) + 1.96\sigma_{\bar{X}}] = .95$$

In terms of the specified numerical values,

$$P[600 - 1.96(16) < \bar{X} < 600 + 1.96(16)] = .95$$
$$P[600 - 31.36 < \bar{X} < 600 + 31.36] = .95$$
$$P[560.64 < \bar{X} < 631.36] = .95$$

If the basic assumptions are true, then the probability that $\bar{X}$ will be between these numbers is indeed .95 and the above statement is a valid probability statement. Thus, a reasonable range for $\bar{X}$ is about 560 to 631. Verification of the assumptions can be made by selecting 25 students and examining the resulting statistics. Suppose it were to happen that $\bar{X} = 619.2$. Since this value of $\bar{X}$ is included in the 95 percent central range, there would be little reason to believe that μ had changed.

11-2 CONFIDENCE INTERVAL FOR μ, ASSUMING σ^2 IS KNOWN

While the result of the previous section is interesting, it isn't very meaningful since the determination of the limits for $\bar{X}$ depends upon knowledge of μ. In general, μ is unknown, and so the previous results are only of academic interest. Furthermore, if μ were known, it would be somewhat foolish and wasteful of time and effort to determine such intervals. Therefore, let the argument of the previous section be turned around and let the discussion begin with the *true* probability statement

$$P[E(\bar{X}) - 1.96\sigma_{\bar{X}} < \bar{X} < E(\bar{X}) + 1.96\sigma_{\bar{X}}] = .95$$

From this probability statement an interval for the unknown parameter μ is easily constructed. The previous probability statement is identical to

$$P\left[-1.96 < \frac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}} < 1.96\right] = .95$$

Within the brackets are two inequalities in which $E(\bar{X})$ is an algebraic unknown. If $\bar{X}$ is known and $\sigma_{\bar{X}}$ is known, then these inequalities can be solved for $E(\bar{X})$. Thus

$$\begin{array}{ll}
-1.96 < \dfrac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}} & \dfrac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}} < 1.96 \\
-1.96\sigma_{\bar{X}} < \bar{X} - E(\bar{X}) & \bar{X} - E(\bar{X}) < 1.96\sigma_{\bar{X}} \\
E(\bar{X}) - 1.96\sigma_{\bar{X}} < \bar{X} & \bar{X} < 1.96\sigma_{\bar{X}} + E(\bar{X}) \\
E(\bar{X}) < \bar{X} + 1.96\sigma_{\bar{X}} & \bar{X} - 1.96\sigma_{\bar{X}} < E(\bar{X})
\end{array}$$

Putting these two results together we see that

$$P[\bar{X} - 1.96\sigma_{\bar{X}} < E(\bar{X}) < \bar{X} + 1.96\sigma_{\bar{X}}] = .95$$

so that the 95 percent confidence interval for $E(\bar{X})$ or μ is given by

$$\bar{X} - 1.96\frac{\sigma}{\sqrt{N}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{N}}$$

This last set of inequalities states that $E(\bar{X})$ is included in an interval that is not dependent upon μ. In other words, the interval can be computed even if μ is unknown. In an equivalent manner, it can be said that $\mu = \bar{X} \pm 1.96\sigma_{\bar{X}}$, where $1.96\sigma_{\bar{X}}$

is the maximum error in the estimate of μ. In this sense, $\pm 1.96\sigma_{\bar{X}}$ measures the precision that $\bar{X}$ has as an estimate of μ.

As an example of the use of confidence intervals in arguing from some to all, consider the following. As part of a nutrition study, 25 boys were selected from an elementary school by a simple random sampling procedure. The height of each boy was taken and then their average height was computed and found to equal 47.23 inches. For boys of this age, national height charts state that the standard deviation in heights is 1.6 inches. If this is the case, then the corresponding 95 percent confidence interval of the expected height of boys of this age at this school is easy to determine. According to the confidence-interval formula, the 95 percent confidence interval is given by

$$\bar{X} - 1.96\frac{\sigma}{\sqrt{N}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{N}}$$

$$47.23 - (1.96)\frac{1.6}{\sqrt{25}} < \mu < 47.23 + (1.96)\frac{1.6}{\sqrt{25}}$$

$$47.23 - .63 < \mu < 47.23 + .63$$

$$46.60 < \mu < 47.86$$

While this is called the 95 percent confidence interval for μ, one should take exceptional care in interpreting the meaning of the interval. Since μ is *not* a random variable but is a *constant* associated with a probability distribution, it is either in the interval 46.60 to 47.86 or it is not. Therefore,

$$P(46.60 < \mu < 47.86) = 0 \text{ or } 1$$

and not .95. While this may sound like a contradiction or paradox, it is easy to clarify the difficulty by performing the following simple experiment.

Take a coin and place it on the thumb, ready to be tossed. Before the coin is tossed, note that the probability it will come up heads is $\frac{1}{2}$, provided that the coin is fair. Toss the coin and when it lands, cover its face before looking at the outcome. Now answer the following question. What is the probability that a head is showing on its upper face? This probability must be 1 or 0, since the coin either came up heads or it did not. One-half is the probability that you can *guess* what the result has been. But since the coin is tossed and the outcome is an event in the past, no probability can be attributed to its outcome except 0 or 1. It either came up heads or it did not.

The same is true of the confidence interval estimation procedure. Either the sample mean generated a confidence interval that includes the true unknown parameter value or it did not. If the resulting interval does contain μ, then the confidence interval represents a true or valid statement. On the other hand, if μ is not included in the interval, then the confidence-interval statement is a false statement. The

resulting interval is called a 95 percent confidence interval because upon *repeated* sampling, 95 percent of the intervals, computed according to the stated procedure of drawing samples of size N and calculating an $\bar{X}$ for each sample, will give confidence intervals that do indeed cover the true value, while 5 percent will miss the mark. Unfortunately, a researcher will never know when a true confidence-interval statement has been made. About all one can do is hope that a computed confidence interval generates a true interval—95 percent of the time it will, 5 percent of the time it will not.

With this introduction, consider the construction of a general confidence interval for μ for any probability level of interest. As a first step to confidence-interval construction, the normal distribution must be algebraically introduced. This can be brought about through

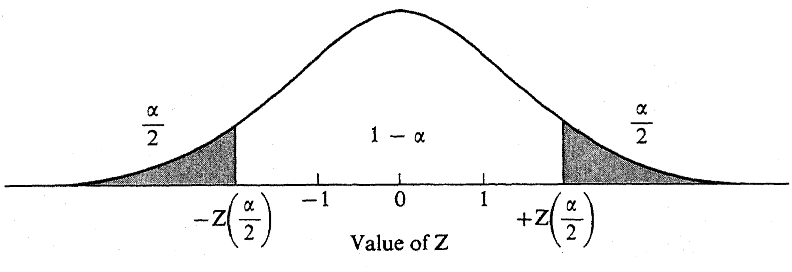
$$Z = \frac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

which is $N(0,1)$, provided that X is normal or N is large enough for the sampling distribution of $\bar{X}$ to come under the influence of the central limit theorem. Since the Z scale extends from minus infinity to plus infinity, it will be necessary to truncate the distribution at two points so as to give an interval of reasonable or usable length. Let these points be denoted as $-Z(\alpha/2)$ and $+Z(\alpha/2)$, as shown in Figure 11-1. This notation is not consistent with the previous notation, which denotes the $\alpha/2$ percentile as $Z(\alpha/2)$ and the $(1-\alpha/2)$ percentile as $Z(1-\alpha/2)$. The only reason for the departure from previous notation is that it makes the resulting formula easier to remember. With this notation,

$$P\left[-Z\left(\frac{\alpha}{2}\right) < Z < Z\left(\frac{\alpha}{2}\right)\right] = 1 - \alpha$$

$$P\left[-Z\left(\frac{\alpha}{2}\right) < \frac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}} < Z\left(\frac{\alpha}{2}\right)\right] = 1 - \alpha$$

Figure 11-1. The normal distribution partitioned into three mutually exclusive subsets with $(1 - \alpha)$ percent of the probability in the central set.



Just as before, the two inequalities can be solved for $E(\bar{X})$ to give

$$P\left[\bar{X} - Z\left(\frac{\alpha}{2}\right)\left(\frac{\sigma_X}{\sqrt{N}}\right) < E(\bar{X}) < \bar{X} + Z\left(\frac{\alpha}{2}\right)\left(\frac{\sigma_X}{\sqrt{N}}\right)\right] = 1 - \alpha$$

so that the $(1 - \alpha)$ percent confidence interval for μ is given by

$$\bar{X} - Z\left(\frac{\alpha}{2}\right)\left(\frac{\sigma_X}{\sqrt{N}}\right) < E(\bar{X}) < \bar{X} + Z\left(\frac{\alpha}{2}\right)\left(\frac{\sigma_X}{\sqrt{N}}\right)$$

When using this inequality to set up an interval estimate for μ, one should check the data to see that the variable of interest has a normal probability distribution, or that N is sufficiently large to ensure coverage under the central limit theorem. If the variable under study is not normally distributed, then a safe rule of thumb states that N should be larger than 30 or thereabouts. It goes without saying that random sampling is needed.

11-3 THE LARGE-SAMPLE CONFIDENCE INTERVAL FOR p, THE PARAMETER OF A BINOMIAL DISTRIBUTION

Exactly the same theory as that used for the confidence interval of μ is also used to establish the confidence interval of p, the parameter of the binomial distribution. As shown earlier, $E(\hat{p}) = p$ and $\text{Var}(\hat{p}) = pq/N$ and if $Np > 5$ and $Nq > 5$, the binomial distribution can be adequately described by the normal distribution. Thus

$$Z = \frac{\hat{p} - E(\hat{p})}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{pq/N}}$$

tends to have a $N(0,1)$ distribution. Therefore

$$P\left[-Z\left(\frac{\alpha}{2}\right) < Z < Z\left(\frac{\alpha}{2}\right)\right] = 1 - \alpha$$

and

$$P\left[-Z\left(\frac{\alpha}{2}\right) < \frac{\hat{p} - E(\hat{p})}{\sigma_{\hat{p}}} < Z\left(\frac{\alpha}{2}\right)\right] = 1 - \alpha$$

With exactly the same algebra as that used for the confidence interval of μ, one can show that

$$P\left[\hat{p} - Z\left(\frac{\alpha}{2}\right)(\sigma_{\hat{p}}) < E(\hat{p}) < \hat{p} + Z\left(\frac{\alpha}{2}\right)(\sigma_{\hat{p}})\right] = 1 - \alpha$$

so that the $(1 - \alpha)$ percent confidence interval for p is given by

$$\hat{p} - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{pq}{N}} < p < \hat{p} + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{pq}{N}}$$

Unfortunately, this formulation is not very useful since p, the unknown parameter, must be known before the confidence interval can be determined. To overcome this difficulty, one simply substitutes $\hat{p}$ for p into the formula for $\sigma_{\hat{p}}$. Upon this substitution, $\sqrt{\hat{p}\hat{q}/N}$ is no longer called the standard deviation of the distribution of $\hat{p}$. Instead, it is called the *standard error* of p and is denoted by $SE_{\hat{p}}$. With this sample estimate, the confidence interval for p becomes

$$\hat{p} - Z\left(\frac{\alpha}{2}\right) SE_{\hat{p}} < p < \hat{p} + Z\left(\frac{\alpha}{2}\right) SE_{\hat{p}}$$

or

$$\hat{p} - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}\hat{q}}{N}} < p < \hat{p} + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}\hat{q}}{N}}$$

As an example of the use of this confidence interval, consider a random sample of 100 children who were asked whether they liked arithmetic. Thirty-six of the children said "Yes." While $\hat{p} = \frac{36}{100} = .36$ is the unbiased efficient point estimate of p, it is clear that if another sample were selected under identical conditions, the results would be different, so that another value of $\hat{p}$ would be obtained. For the obtained sample, $N\hat{p} = 36$ and $N\hat{q} = 64$ are both greater than 5. As a result, the normal approximation is probably appropriate and a reasonable 95 percent interval for p is given by

$$\hat{p} - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}\hat{q}}{N}} < p < \hat{p} + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}\hat{q}}{N}}$$

$$.36 - 1.96\sqrt{\frac{(.36)(.64)}{100}} < p < .36 + 1.96\sqrt{\frac{(.36)(.64)}{100}}$$

$$.36 - .09 < p < .36 + .09$$

$$.27 < p < .45$$

There appears to be little doubt that less than 50 percent of the children report that they like arithmetic. Again, the probability that the true proportion of students who say they like arithmetic is included in the interval from .27 to .45 is either 1 or 0. Either the unknown p is in the interval or it is not. The probability that p is in this interval is definitely not .95.

Whereas the assumptions required for the determination of the confidence interval of μ were not emphasized, one must be a little more concerned about the assumptions that are used to define the confidence interval for p. First, independence of selection and independence of response by the children is a necessity. Furthermore, constant probability of a "yes" response for all children who select that response is required for the variable to have a binomial distribution. While the first assumption is not unreasonable, the second is questionable. The degree to which arithmetic appeals to

children varies with the child. As an educator would say, there are individual differences across students with respect to their attitudes toward arithmetic. There are many children who like arithmetic, and for them, p is large. For those children who dislike arithmetic, p is low. Thus, in the population of children one would expect to find a population of $\hat{p}$'s. While this may appear to mitigate against the binomial assumption, the behavioral scientist need not be overly concerned, for even though this assumption is clearly violated, the confidence-interval method tends to give interval limits for the average population p that are useful and perhaps not far from wrong for the average p.

One other thing that must be checked for when setting up a confidence interval for an unknown p is whether $N\hat{p} > 5$ and $N\hat{q} > 5$. Unless this condition is satisfied, one should hesitate to use the confidence-interval formula of this section since it is based upon the approximation of the binomial distribution to the normal distribution.

Finally, it should be noted that for small sample sizes a correction for continuity is advisable since the discrete property of a binomial variable can produce a sizable error. For the small-sample case, one is advised to use

$$\left(\hat{p} - \frac{1}{2N}\right) - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}\hat{q}}{N}} < p < \left(\hat{p} + \frac{1}{2N}\right) + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}\hat{q}}{N}}$$

Naturally, this correction is also appropriate for large samples or for any size sample.

11-4 CONFIDENCE INTERVALS FOR μ WHEN σ^2 IS UNKNOWN

As was shown in Section 10-5, $(\bar{X} - \mu)/(S_X/\sqrt{N})$ has a t distribution with $\nu = N - 1$ degrees of freedom. Following the procedure employed for the two preceding confidence intervals, one can choose two numbers, $-t_\nu(\alpha/2)$ and $+t_\nu(\alpha/2)$, so that

$$P\left[-t_\nu\left(\frac{\alpha}{2}\right) < t < t_\nu\left(\frac{\alpha}{2}\right)\right] = 1 - \alpha$$

Substituting

$$t = \frac{\bar{X} - \mu}{S_X/\sqrt{N}}$$

into this set of inequalities, we find that:

$$P\left[-t_\nu\left(\frac{\alpha}{2}\right) < \frac{\bar{X} - \mu}{S_X/\sqrt{N}} < t_\nu\left(\frac{\alpha}{2}\right)\right] = 1 - \alpha$$

As before, these inequalities can be solved for μ to give

$$P\left[\bar{X} - t_\nu\left(\frac{\alpha}{2}\right)\left(\frac{S_X}{\sqrt{N}}\right) < \mu < \bar{X} < t_\nu\left(\frac{\alpha}{2}\right)\left(\frac{S_X}{\sqrt{N}}\right)\right] = 1 - \alpha$$

so that the $(1 - \alpha)$ percent confidence interval for μ is given by

$$\bar{X} - t_\nu\left(\frac{\alpha}{2}\right)\left(\frac{S_X}{\sqrt{N}}\right) < \mu < \bar{X} < t_\nu\left(\frac{\alpha}{2}\right)\left(\frac{S_X}{\sqrt{N}}\right)$$

As an example of the use of this result, consider a random sample of 20 second-year students majoring in German who were given a difficult passage of scientific German to read. The time taken to read the material was determined for each student. The sample statistics were as follows:

$\bar{X} = 167$ seconds and $S_X = 42$ seconds

The degrees of freedom used to estimate S_X is $\nu = N - 1 = 20 - 1 = 19$. Thus,

$t_{19}(.025) = -2.093$ and $t_{19}(.975) = 2.093$

The 95 percent confidence interval for μ is given by

$$\bar{X} - t_\nu\left(\frac{\alpha}{2}\right)\left(\frac{S_X}{\sqrt{N}}\right) < \mu < \bar{X} + t_\nu\left(\frac{\alpha}{2}\right)\left(\frac{S_X}{\sqrt{N}}\right)$$

$$167 - 2.093\left(\frac{42}{\sqrt{20}}\right) < \mu < 167 + 2.093\left(\frac{42}{\sqrt{20}}\right)$$

$$167 - 19.66 < \mu < 167 + 19.66$$

$$147.34 < \mu < 186.66$$

From this it can be concluded that the average time taken to read the passage is anywhere from about 147 seconds to about 187 seconds.

The assumptions required for the determination of this confidence interval are random sampling and normality of the observed variable. The assumption of normality is of some significance since reading times tend to be positively skewed. If the distribution is extremely pathological, then the normality assumption is important if N is small. As N increases, the normality assumption can be weakened. For most of the variables encountered in the behavioral sciences, the normality assumption may not be too critical. Since the t distribution is insensitive to minor departures from normality, the behavioral scientist need only check to see that the deviation from normality in the sample distribution is not severe.

11-5 THE STANDARD ERROR OF A STATISTIC

The standard deviation of the sampling distribution of means is given by

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{N}}$$

This standard deviation appears in the denominator of

$$Z = \frac{\bar{X} - E(\bar{X})}{\sigma_X/\sqrt{N}}$$

and in the confidence interval for $E(X)$:

$$\bar{X} - Z\left(\frac{\alpha}{2}\right)\left(\frac{\sigma_X}{\sqrt{N}}\right) < E(X) < \bar{X} + Z\left(\frac{\alpha}{2}\right)\left(\frac{\sigma_X}{\sqrt{N}}\right)$$

When the standard deviation of the population is unknown, the standard deviation of the sampling distribution of $\bar{X}$ is also unknown. In this case, $S_X/\sqrt{N}$ can be used as an estimate of $\sigma_X/\sqrt{N}$. This estimate is called the standard error of the mean. The standard error of $\bar{X}$ appears in the denominator of

$$t = \frac{\bar{X} - E(\bar{X})}{S_X/\sqrt{N}}$$

and in the confidence interval for $E(X)$:

$$\bar{X} - t\left(\frac{\alpha}{2}\right)\left(\frac{S_X}{\sqrt{N}}\right) < E(X) < \bar{X} + t\left(\frac{\alpha}{2}\right)\left(\frac{S_X}{\sqrt{N}}\right)$$

As shown in Section 9-10, the standard deviation of the sampling distribution of $\hat{p}$ is given by $\sigma_{\hat{p}} = \sqrt{pq/N}$. This formula appears in the normal approximation formula

$$Z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{pq/N}}$$

but note that the standard error of $\hat{p}$, that is, $\mathrm{SE}_{\hat{p}} = \sqrt{\hat{p}\hat{q}/N}$, appears in the confidence interval statement for p:

$$\hat{p} - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}\hat{q}}{N}} < p < \hat{p} + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}\hat{q}}{N}}$$

11-6 CONFIDENCE INTERVAL FOR A GENERAL PARAMETER

Confidence intervals for μ and p have a similar form, and as might be suspected, both can be derived from a general formula that depends upon the large sample characteristics of sampling distributions of parameter estimators. For a general model, let θ be a parameter to be estimated. Let $\hat{\theta}$ be an estimator for θ, which upon repeated sampling has normal distribution. Operationally, this means that the sampling distribution of $\hat{\theta}$ is approximately normal if N is relatively large. Let $E(\hat{\theta}) = \theta$ and let the variance of the sampling distribution of $\hat{\theta}$ be denoted by $\sigma^2_{\hat{\theta}}$.

Under these conditions, $Z = (\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$ is approximately $N(0,1)$. The $(1 - \alpha)$ percent central region of this distribution is given by

$$-Z\left(\frac{\alpha}{2}\right) < \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} < Z\left(\frac{\alpha}{2}\right)$$

Solving these inequalities for θ, we find that the $(1 - \alpha)$ percent confidence interval for θ is given by

$$\hat{\theta} - Z\left(\frac{\alpha}{2}\right)\sigma_{\hat{\theta}} < \theta < \hat{\theta} + Z\left(\frac{\alpha}{2}\right)\sigma_{\hat{\theta}}$$

If $\sigma_{\hat{\theta}}$ is unknown, then one can substitute its estimate, $\hat{\sigma}_{\hat{\theta}} = \mathrm{SE}_{\hat{\theta}}$, into the formula and replace Z by t to obtain the $(1 - \alpha)$ percent confidence interval for θ. For this model, the $(1 - \alpha)$ percent confidence interval for θ is given by

$$\hat{\theta} - t_{\nu}\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\theta}} < \theta < \hat{\theta} + t_{\nu}\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\theta}}$$

For both models, random samples are required and N must be large enough so that the central limit theorem takes effect in inducing normality or approximate normality of the sampling distribution of $\hat{\theta}$.

11-7 CONFIDENCE INTERVAL FOR THE DIFFERENCES OF TWO EXPECTED VALUES (VARIANCES KNOWN)

The most frequent research design of behavioral research involves the comparing of two probability distributions. In its most idealized form, the basic research question involves the comparison of an experimental group with a control group. The appropriate statistical procedures for this idealized experimental model are presented in this section.

Let X_1 be $N(\mu_1,\sigma_1^2)$ and let X_2 be $N(\mu_2,\sigma_2^2)$. Let independent random samples of size N_1 and N_2 be selected from each population, respectively. Let $\theta = \mu_1 - \mu_2$. According to this model, θ measures the average difference in the expected values of the two populations, as is shown in the following two theorems. An unbiased estimate of θ is given by $\hat{\theta} = \bar{X}_1 - \bar{X}_2$, with $\sigma_{\hat{\theta}}^2 = \sigma_1^2/N_1 + \sigma_2^2/N_2$.

Theorem 11-1

The expected value of the difference of two means is equal to the difference in their expected values.

Proof. By definition and Theorem 6-1,

$$\begin{aligned}
E(\hat{\theta}) &= E(\bar{X}_1 - \bar{X}_2) \\
&= E[(1)\bar{X}_1 + (-1)\bar{X}_2] \\
&= E[(+1)\bar{X}_1] + E[(-1)\bar{X}_2] \\
&= (+1)E(\bar{X}_1) + (-1)E(\bar{X}_2) \\
&= E(\bar{X}_1) - E(\bar{X}_2) \\
&= \mu_1 - \mu_2 \\
&= \theta
\end{aligned}$$

This completes the proof.

Theorem 11-2

The variance of the sampling distribution of $\hat{\theta} = \bar{X}_1 - \bar{X}_2$ is given by

$$\sigma_{\hat{\theta}}^2 = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}$$

Proof. By definition and Theorem 6-4,

$$\begin{aligned} \sigma_{\hat{\theta}}^2 = \operatorname{Var}(\hat{\theta}) &= \operatorname{Var}(\bar{X}_1 - \bar{X}_2) \\ &= \operatorname{Var}[(+1)\bar{X}_1 + (-1)\bar{X}_2] \\ &= (+1)^2 \operatorname{Var}(\bar{X}_1) + (-1)^2 \operatorname{Var}(\bar{X}_2) \\ &= \operatorname{Var}(\bar{X}_1) + \operatorname{Var}(\bar{X}_2) \\ &= \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} \end{aligned}$$

This completes the proof.

With these results, the $(1 - \alpha)$ percent confidence interval for θ is given by

$$(\bar{X}_1 - \bar{X}_2) - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

As an example of the use of this interval, consider $N_B = 85$ boys and $N_G = 72$ girls who were given a standardized test for which $\sigma_B = 18$ and $\sigma_G = 20$. The average for the boys was $\bar{X}_B = 36$ and the average for the girls was $\bar{X}_G = 51$. The 95 percent confidence interval for $\theta = \mu_B - \mu_G$ is given by

$$(\bar{X}_B - \bar{X}_G) - 1.96\sqrt{\frac{\sigma_B^2}{N_B} + \frac{\sigma_G^2}{N_G}} < \mu_B - \mu_G < (\bar{X}_B - \bar{X}_G) + 1.96\sqrt{\frac{\sigma_B^2}{N_B} + \frac{\sigma_G^2}{N_G}}$$

$$(36 - 51) - 1.96\sqrt{\frac{18^2}{85} + \frac{20^2}{72}} < \mu_B - \mu_G < (36 - 51) + 1.96\sqrt{\frac{18^2}{85} + \frac{20^2}{72}}$$

$$-15 - 1.96\sqrt{3.81176 + 5.5556} < \mu_B - \mu_G < -15 + 1.96\sqrt{3.81176 + 5.5556}$$

$$-15 - 1.96\sqrt{9.36732} < \mu_B - \mu_G < -15 + 1.96\sqrt{9.36732}$$

$$-15 - 1.96(3.0606) < \mu_B - \mu_G < -15 + 1.96(3.0606)$$

$$-15 - 6.00 < \mu_B - \mu_G < -15 + 6.00$$

$$-21.00 < \mu_B - \mu_G < -9.00$$

This result suggests that the girls out-perform the boys by about 21 to about 9 test-score points.

11-8 CONFIDENCE INTERVAL FOR THE DIFFERENCE OF TWO EXPECTED VALUES WHEN VARIANCES ARE UNKNOWN AND EQUAL

The model of the previous section is somewhat artificial, since it was assumed that the variances were known. In general, they are unknown. As a special, and yet frequent, model of behavioral research, suppose that the variances are unknown but equal in value. As shown in Section 11-7, an unbiased estimate of $\theta = \mu_1 - \mu_2$ is given by $\hat{\theta} = \bar{X}_1 - \bar{X}_2$. Also, as shown, $\text{Var}(\hat{\theta}) = \sigma_1^2/N_1 + \sigma_2^2/N_2$, but since it has been assumed that $\sigma_1^2 = \sigma_2^2 = \sigma_0^2$, it follows that

$$\text{Var}(\hat{\theta}) = \frac{\sigma_0^2}{N_1} + \frac{\sigma_0^2}{N_2} = \sigma_0^2\left(\frac{1}{N_1} + \frac{1}{N_2}\right)$$

However, σ_0^2 is unknown and must be estimated. The resulting estimate of the $\text{Var}(\hat{\theta})$ is given by

$$\text{SE}_{\hat{\theta}}^2 = S_p^2\left(\frac{1}{N_1} + \frac{1}{N_2}\right)$$

where S_p^2 is a pooled sample estimate of σ_0^2.

As shown earlier, S_1^2 is an unbiased estimate of σ_0^2 and S_2^2 is also an unbiased estimate of σ_0^2. There are a number of ways they can be put together to give a good estimate. One of the best ways is to compute the deviation of each observation from its own mean and then use the total set of squared deviations to estimate σ_0^2. For the first sample, $(X_{11}, X_{12}, \ldots, X_{1N_1})$, the deviations are given by $(X_{11} - \bar{X}_1, X_{12} - \bar{X}_1, \ldots, X_{1N_1} - \bar{X}_1)$, so that the sum of the squared deviations for the first sample is given by

$$\text{SS}_1 = \sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2$$

In an exactly analogous manner, the sum of the squared deviations for the second sample is given by

$$\text{SS}_2 = \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2$$

For both samples, the total sum of the squared deviations is given by

$$\text{SS}_1 + \text{SS}_2 = \sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2$$

In the first sample, $\nu_1 = (N_1 - 1)$, while in the second sample, $\nu_2 = (N_2 - 1)$. Thus, S_p^2 is based on $\nu_1 + \nu_2 = (N_1 - 1) + (N_2 - 1)$ degrees of freedom. As a result,

$$S_p^2 = \frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{(N_1 - 1) + (N_2 - 1)}$$

Since

$$S_1^2 = \frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1 - 1}$$

it follows that

$$(N_1 - 1)\, S_1^2 = \sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2$$

For the second sample, one has

$$(N_2 - 1)\, S_2^2 = \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2$$

Substituting these values into S_p^2, we find that

$$S_p^2 = \frac{(N_1 - 1)\, S_1^2 + (N_2 - 1)\, S_2^2}{(N_1 - 1) + (N_2 - 1)}$$

In this form it is seen that S_p^2 is a weighted average of the sample variances where the weights are the respective degrees of freedom. It is worth noting that this estimate is based upon the variability within each of the two samples. For this reason it is called the "within sample estimate of the common unknown variance." This estimate is unbiased, as will now be shown.

Theorem 11-3

$$E(S_p^2) = \sigma_0^2$$

Proof. By definition and Theorem 6-1,

$$E(S_p^2) = E\left[\frac{(N_1 - 1)\, S_1^2}{N_1 + N_2 - 2} + \frac{(N_2 - 1)\, S_2^2}{N_1 + N_2 - 2}\right]$$

$$= E\left[\frac{(N_1 - 1)\, S_1^2}{N_1 + N_2 - 2}\right] + E\left[\frac{(N_2 - 1)\, S_2^2}{N_1 + N_2 - 2}\right]$$

Since the sample sizes are constants and not variables, it follows by Theorem 7-3 that

$$E(S_p^2) = \frac{N_1 - 1}{N_1 + N_2 - 2} E(S_1^2) + \frac{N_2 - 1}{N_1 + N_2 - 2} E(S_2^2)$$

By assumption, $E(S_1^2) = E(S_2^2) = \sigma_0^2$, so that

$$E(S_p^2) = \frac{(N_1 - 1)\, \sigma_0^2}{N_1 + N_2 - 2} + \frac{(N_2 - 1)\, \sigma_0^2}{N_1 + N_2 - 2}$$

$$= \sigma_0^2$$

This completes the proof.

The estimate S_p^2 is based upon $\nu = (N_1 - 1) + (N_2 - 1) = N_1 + N_2 - 2$ degrees of freedom. Thus, the $(1 - \alpha)$ percent confidence interval for $\theta = \mu_1 - \mu_2$ is given by

$$(\bar{X}_1 - \bar{X}_2) - t_\nu\left(\frac{\alpha}{2}\right) S_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_\nu\left(\frac{\alpha}{2}\right) S_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

As an example of the use of this confidence interval, consider the following study. In a curriculum study designed to select an efficient method for teaching arithmetic to third-grade children, 23 children in a culturally deprived neighborhood were randomly assigned to a control and an experimental teaching condition. The control group was taught with traditional methods while the experimental group was taught with new methods that employed multicolored teaching devices of interesting form. At the end of two weeks of teaching, both groups were given the same achievement test. The results were as reported in Table 11-1. For these observations,

Table 11-1. Scores made by 23 children in a recent curriculum study on arithmetic teaching.

Control method	*Experimental method*
20	32
31	41
17	39
29	32
32	33
39	47
16	35
42	48
27	31
30	31
25	29
	22

$$\bar{X}_C = \frac{1}{N_C}\sum_{i=1}^{11} X_{Ci} = \frac{308}{11} = 28.0 \qquad \bar{X}_E = \frac{1}{N_E}\sum_{i=1}^{12} X_{Ei} = \frac{420}{12} = 35.0$$

so that an unbiased estimate of $\theta = \mu_C - \mu_E$ is given by

$$\hat{\theta} = \bar{X}_C - \bar{X}_E = 28 - 35 = -7$$

For these data,

$$S_C^2 = \frac{N_C \sum X_{Ci}^2 - (\sum X_{Ci})^2}{N_C(N_C - 1)} \qquad S_E^2 = \frac{N_E \sum X_{Ei}^2 - (\sum X_{Ei})^2}{N_E(N_E - 1)}$$

$$= \frac{11(9310) - (308)^2}{11(10)} \qquad = \frac{12(15324) - (420)^2}{12(11)}$$

$$= 68.60 \qquad = 56.73$$

While the separate sample variances are not equal, they probably are within sampling errors of one another so that the assumption of common variance is not unrealistic. Therefore, they are combined to estimate the common value. Thus

$$S_p^2 = \frac{(N_C - 1)\, S_C^2 + (N_E - 1)\, S_E^2}{(N_C - 1) + (N_E - 1)}$$

$$= \frac{(11 - 1)(68.60) + (12 - 1)(56.73)}{(11 - 1) + (12 - 1)}$$

$$= 62.38$$

This estimate is based upon

$$\nu = (N_C - 1) + (N_E - 1) = (11 - 1) + (12 - 1) = 21$$

degrees of freedom.

For this value of ν,

$$t_{21}(.025) = -2.080 \qquad \text{and} \qquad t_{21}(.975) = 2.080$$

so that the 95 percent confidence interval for $\theta = \mu_C - \mu_E$ is given by

$$-7 - 2.080\sqrt{\frac{62.38}{11} + \frac{62.38}{12}} < \mu_C - \mu_E < -7 + 2.080\sqrt{\frac{62.38}{11} + \frac{62.38}{12}}$$

$$-7 - 2.080\sqrt{10.87} < \mu_C - \mu_E < -7 + 2.080\sqrt{10.87}$$

$$-7 - 6.86 < \mu_C - \mu_E < -7 + 6.86$$

$$-13.86 < \mu_C - \mu_E < -.14$$

Note that $\mu_C - \mu_E = 0$ is not included in this interval. Therefore, it makes sense to conclude that the experimental method is better than the control method even though the average difference may not exceed .14 of a point.

The assumptions employed in this analysis are:

1. Random samples have been selected from the control and experimental populations.
2. The observations on performance in the tests between and within samples are statistically independent.

3. In the two hypothetical populations of scores, the variances are equal.
4. Either the parent score populations are normal or else the samples are large enough to generate sampling distributions of means that are close enough to normality for the probabilities to be valid.

In this case, these assumptions are not unreasonable

It should be emphasized that the populations of the example are hypothetical in nature. The control population is an artificial population that is generated by the random assignment of students to the control method of teaching. In an abstract sense there is a population of students that could have taken part in this study. However, it is impossible to make a list or roster of the students who comprise this population. The population exists only in a hypothetical sense. This is also true of the population defined by the experimental method. This dependence on inference procedures from samples to hypothetical populations is not peculiar to this example. It is a characteristic of most behavioral research populations.

11-9 CONFIDENCE INTERVALS FOR THE DIFFERENCE OF TWO EXPECTED VALUES WHEN THE VARIANCES ARE UNKNOWN AND NOT NECESSARILY EQUAL

According to the general theory, a confidence interval for $\theta = \mu_1 - \mu_2$ is given by

$$(\bar{X}_1 - \bar{X}_2) - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

When the variances are known, this interval can be determined. When the variances are unknown and equal the interval is given by

$$(\bar{X}_1 - \bar{X}_2) - t\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_p^2}{N_1} + \frac{S_p^2}{N_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_p^2}{N_1} + \frac{S_p^2}{N_2}}$$

When the variances are unknown and unequal, the problem is more difficult and exact solutions are not available. A commonly used solution to the problem was proposed by Welch and Aspin, 1949, and their solution is as follows.

In place of $\sigma_{\hat{\theta}} = \sqrt{\sigma_1^2/N_1 + \sigma_2^2/N_2}$, one simply substitutes the standard error of the difference of two averages into the first of the two formulas and then uses the t distribution in place of Z. With these substitutions,

$$\mathrm{SE}_{\hat{\theta}} = \mathrm{SE}_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}$$

The confidence interval for θ is given by

$$(\bar{X}_1 - \bar{X}_2) - t_{\nu*}^*\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\nu*}^*\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}$$

With this formulation, every parameter is estimable from the samples and as a result, one could determine the limits of the confidence interval provided that the degrees of freedom of t^* are known.

Welch and Aspin have shown that the degrees of freedom can be approximated by the following formula:

$$\nu^* = \frac{\nu_1 \nu_2}{\nu_2 C^2 + \nu_1(1 - C)^2}$$

where

$$C = \frac{SE^2_{\bar{X}_1}}{SE^2_{\bar{X}_1} + SE^2_{\bar{X}_2}}$$

As an illustration of the use of this result, reconsider the determination of the 95 percent confidence interval for the difference in expected value for the previously discussed example. In that example,

1. $\hat{\theta} = \bar{X}_1 - \bar{X}_2 = 28 - 35 = -7$
2. $\nu_1 = N_1 - 1 = 11 - 1 = 10$
3. $\nu_2 = N_2 - 1 = 12 - 1 = 11$
4. $SE^2_{\bar{X}_1} = \dfrac{S_1^2}{N_1} = \dfrac{68.60}{11} = 6.24$
5. $SE^2_{\bar{X}_2} = \dfrac{S_2^2}{N_2} = \dfrac{56.73}{12} = 4.73$
6. $SE^2_{\bar{X}_1} + SE^2_{\bar{X}_2} = 6.24 + 4.73 = 10.97$
7. $C = \dfrac{6.24}{10.97} = .5688$
8. $$\begin{aligned}\nu^* &= \frac{(10)(11)}{11(.5688)^2 + 10(.4312)^2} \\ &= \frac{110}{3.5588 + 1.8593} \\ &= \frac{110}{5.41810} \\ &= 20.30 \\ &\cong 20\end{aligned}$$
9. $t^*_{20}(.025) = -2.086$

The 95 percent confidence interval is given by

$$-7 - 2.086\sqrt{10.97} < \mu_C - \mu_E < -7 + 2.086\sqrt{10.97}$$
$$-7 - 6.91 < \mu_C - \mu_E < -7 + 6.91$$
$$-13.91 < \mu_C - \mu_E < -.09$$

so that exactly the same conclusions are made.

In this case, the degrees of freedom for t^* have been reduced from 21 to 20. The reason for the small reduction in degrees of freedom is directly attributable to the near equality of the sample variances. If the sample variances had been very different from one another, the reduction would have been greater.

Fortunately, one need not compute the degrees of freedom for every case in which unequal variances exist. The reason for this will now be explained.

Let ν_1 = degrees of freedom for sample 1, ν_2 = degrees of freedom for sample 2, and $\nu = \min(\nu_1, \nu_2)$. It can be shown that $\nu^* \geqslant \nu$. Thus, the degrees of freedom for the Welch–Aspin intervals will always be greater than or equal to the degrees of freedom for the smaller sample. For example, if $\nu_1 = 38$ and $\nu_2 = 84$, then $\nu^* \geqslant 38$. Thus, if $N_1 > 30$ and $N_2 > 30$, one can immediately use the normal approximation. It is only when $N_1 < 30$ or $N_2 < 30$ that one must compute the Welch–Aspin degrees of freedom.

11-10 CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO EXPECTED VALUES WHEN THE SAMPLES ARE NOT STATISTICALLY INDEPENDENT (COMMON VARIANCE NOT ASSUMED)

All the confidence-interval procedures presented for the difference in expected values of two populations have depended upon the assumption that the samples have been selected independently of one another. Many experimental situations exist where this does not happen, and a frequently encountered example of this is in a study that employs pre- and posttests to achieve some sort of experimental control.

For example, four students were selected to serve as subjects in a reading experiment. Their reading rates were determined prior to the experiment. Following the experiment, their reading rates were again determined. The results of the study are shown in Table 11-2.

Table 11-2. Words per minute read by four students participating in a reading study designed to increase reading rates.

Student	*Posttest score*	*Pretest score*	*Difference score*
Alice	312	183	129
Bob	293	245	48
Carol	322	202	120
Donald	261	178	83

One of the desirable features of this study is that each student serves as his own control. The criterion variable of interest is the difference score, or improvement score in reading rate.

If X_1 = score on the posttest and X_2 = score on the pretest, then the parameter of interest is given by

$$\begin{aligned}\theta &= E(X_1) - E(X_2) \\ &= \mu_1 - \mu_2 \\ &= \mu_d\end{aligned}$$

If one computes the differences for each subject so that $d_i = X_{1i} - X_{2i}$, and if the difference scores are normally distributed, a confidence interval for μ_d is given simply as

$$\bar{d} - t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\bar{d}} < \mu_d < \bar{d} + t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\bar{d}}$$

For this example one should pay attention to the normality assumption since reading rates tend to be positively skewed. For large samples, the normality assumption can be relaxed since $\bar{d}$ will tend to a normal form because of the central limit theorem. If one assumes that the normality assumption is reasonably satisfied, then

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i = \frac{129 + 48 + 120 + 83}{4} = \frac{380}{4} = 95$$

$$\begin{aligned}S_d^2 &= \frac{N\left(\sum_{i=1}^{N} d_i^2\right) - \left(\sum_{i=1}^{N} d_i\right)^2}{N(N-1)} \\ &= \frac{4(129^2 + 48^2 + 120^2 + 83^2) - 380^2}{4(3)} \\ &= 1378\end{aligned}$$

and

$$\mathrm{SE}_{\bar{d}} = \sqrt{\frac{S_d^2}{N}} = \sqrt{\frac{1378}{4}} = 18.56$$

Since $N = 4$, $\nu = N - 1 = 3$, and $t_3(.025) = -3.182$, the 95 percent confidence interval for μ_d is given by

$$\begin{aligned}\bar{d} - t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\bar{d}} &< \mu_d < \bar{d} + t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\bar{d}} \\ 95 - 3.182(18.56) &< \mu_d < 95 + 3.182(18.56) \\ 95 - 59.06 &< \mu_d < 95 + 59.06 \\ 35.94 &< \mu_d < 154.06\end{aligned}$$

According to these results, the minimum average improvement is expected to be about 36 words per minute, and the maximum average improvement is about 154 words per minute. As this example shows, one must always check to see whether the samples are independent. If they are not, then this procedure should be used.

11-11 CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO BINOMIAL PARAMETERS

Not all comparative studies use continuous or discrete variables as criterion measures. Frequently, the criterion variables are qualitative or dichotomous. When this occurs, confidence intervals for the difference in percentage values are desired. For these types of studies, let X_1 be $B(N_1, p_1)$ and let X_2 be $B(N_2, p_2)$. Let $\theta = p_1 - p_2$. An unbiased estimate of θ is given by $\hat{\theta} = \hat{p}_1 - \hat{p}_2$ as is now shown.

Theorem 11-4

$\hat{\theta} = \hat{p}_1 - \hat{p}_2$ is an unbiased estimate of $\theta = p_1 - p_2$.

Proof. By definition and Theorem 6-1,

$$\begin{aligned} E(\hat{\theta}) &= E(\hat{p}_1 - \hat{p}_2) \\ &= E(\hat{p}_1) - E(\hat{p}_2) \\ &= p_1 - p_2 \\ &= \theta \end{aligned}$$

This completes the proof.

The variance of θ is given by $\operatorname{Var}(\hat{\theta}) = p_1 q_1 / N_1 + p_2 q_2 / N_2$, as is now shown.

Theorem 11-5

$$\operatorname{Var}(\hat{\theta}) = \operatorname{Var}(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2}$$

Proof. By definition and Theorem 6-3,

$$\begin{aligned} \operatorname{Var}(\hat{\theta}) &= \operatorname{Var}(\hat{p}_1 - \hat{p}_2) \\ &= \operatorname{Var}(\hat{p}_1) + \operatorname{Var}(\hat{p}_2) \\ &= \frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2} \end{aligned}$$

This completes the proof.

Since p_1 and p_2 are unknown, the $\mathrm{Var}(\hat{\theta})$ cannot be computed. Therefore, one must substitute the sample estimates into $\sigma_{\hat{\theta}}^2$ and use the standard error of the difference:

$$\mathrm{SE}_{\hat{\theta}} = \mathrm{SE}_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1} + \frac{\hat{p}_2\hat{q}_2}{N_2}}$$

so that the $(1-\alpha)$ percent confidence interval for θ is given by

$$(\hat{p}_1 - \hat{p}_2) - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1} + \frac{\hat{p}_2\hat{q}_2}{N_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1} + \frac{\hat{p}_2\hat{q}_2}{N_2}}$$

As with all binomial approximations to the normal distribution, one needs to ensure that

$$N_1 p_1 > 5 \qquad N_2 p_2 > 5 \qquad N_1 q_1 > 5 \qquad \text{and} \qquad N_2 q_2 > 5$$

Since the parameters are unknown, one should check to see that

$$N_1 \hat{p}_1 > 5 \qquad N_2 \hat{p}_2 > 5 \qquad N_1 \hat{q}_1 > 5 \qquad \text{and} \qquad N_2 \hat{q}_2 > 5$$

Furthermore, it must be assumed that the random samples from each of the populations are statistically independent.

As a use of this result, consider the following example. In a study of the attitudes of the adults in Berkeley, California toward the integration of their schools, members of a random sample were asked: "For some elementary schools, it has been suggested that lines be changed so that the percentage of nonwhite and white children in these schools would be more like the percentage for the entire school system." The results are shown in Table 11-3. In light of the discussion on statistical indepen-

Table 11-3. Responses by race to a question asked of adults in Berkeley, California about changing school boundaries.

	White	*Nonwhite*	*Total*
Agree	297	153	450
Disagree	281	30	311
Total	578	183	761

dence in Section 3-10, it is clear that the question of the study relates to whether or not attitudes are independent of race. In this case, it is doubtful, since

$$\hat{p}_W = P(A \mid W) = \tfrac{297}{578} = .514$$
$$\hat{p}_{\bar{W}} = P(A \mid \bar{W}) = \tfrac{153}{183} = .836$$
$$\hat{p}_0 = P(A) = \tfrac{450}{761} = .591$$

Instead of using the rule of thumb presented in Section 3-10, one can determine whether or not the attitudes are independent of race by examining the confidence interval. For this case,

$$\hat{\theta} = \hat{p}_W - \hat{p}_{\bar{W}} = .514 - .836 = -.322$$

$$\begin{aligned} SE_{\hat{\theta}}^2 &= \frac{\hat{p}_W \hat{q}_W}{N_W} + \frac{\hat{p}_{\bar{W}} \hat{q}_{\bar{W}}}{N_{\bar{W}}} \\ &= \frac{(.514)(.486)}{578} + \frac{(.836)(.164)}{183} \\ &= .00041 + .00074 \\ &= .00115 \\ SE_{\hat{\theta}} &= \sqrt{.00115} \\ &= .0339 \end{aligned}$$

so that the 95 percent confidence interval is given by

$$\hat{\theta} - Z\left(\frac{\alpha}{2}\right) SE_{\hat{\theta}} < \theta < \hat{\theta} + Z\left(\frac{\alpha}{2}\right) SE_{\hat{\theta}}$$

$$-.322 - 1.96(.0339) < \theta < -.322 + 1.96(.0339)$$

$$-.322 - .066 < \theta < -.322 + .066$$

$$-.388 < p_W - p_{\bar{W}} < -.256$$

As suggested now by the exclusion of 0, the attitudes of whites and nonwhites toward the integration of the two schools are quite different. A reasonable estimate of the minimum difference in attitudes is given by .256. This suggests that whatever the percent agreement for whites may be, the percent agreement for nonwhites is higher by at least 26 percent.

11-12 CONFIDENCE INTERVAL FOR σ^2 OF THE NORMAL DISTRIBUTION

As was shown in Section 10-4, when X is normal with $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$, then

$$\chi_\nu^2 = \frac{(N-1)S^2}{\sigma^2}$$

has a chi-square distribution with $\nu = (N-1)$ degrees of freedom. Since this equation involves both S^2 and σ^2, one might think that the chi-square distribution could be

used to determine a confidence interval for σ^2, and indeed it can. Selecting two points from a chi-square distribution, as shown in Figure 11-2, we have

$$P\left[\chi^2_{N-1}\left(\frac{\alpha}{2}\right) < \chi^2 < \chi^2_{N-1}\left(1-\frac{\alpha}{2}\right)\right] = 1-\alpha$$

where $\chi^2_{N-1}(\alpha/2)$ is the $\alpha/2$ percentile of the chi-square distribution with $(N-1)$ degrees of freedom, and $\chi^2_{N-1}(1-\alpha/2)$ is the $(1-\alpha/2)$ percentile. Substituting $(N-1)S^2/\sigma^2$ for χ^2, we have

$$P\left[\chi^2_{N-1}\left(\frac{\alpha}{2}\right) < \frac{(N-1)S^2}{\sigma^2} < \chi^2_{N-1}\left(1-\frac{\alpha}{2}\right)\right] = 1-\alpha$$

As before, there are two inequalities that must be solved for σ^2. These inequalities and their solutions are given by

$$\chi^2_{N-1}\left(\frac{\alpha}{2}\right) < \frac{(N-1)S^2}{\sigma^2} \qquad \frac{(N-1)S^2}{\sigma^2} < \chi^2_{N-1}\left(1-\frac{\alpha}{2}\right)$$

$$\sigma^2\chi^2_{N-1}\left(\frac{\alpha}{2}\right) < (N-1)S^2 \qquad (N-1)S^2 < \sigma^2\chi^2_{N-1}\left(1-\frac{\alpha}{2}\right)$$

$$\sigma^2 < \frac{(N-1)S^2}{\chi^2_{N-1}(\alpha/2)} \qquad \frac{(N-1)S^2}{\chi^2_{N-1}(1-\alpha/2)} < \sigma^2$$

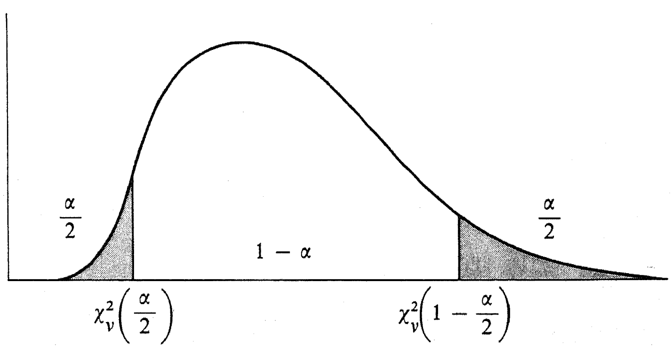
Putting these two inequalities together, we have

$$P\left[\frac{(N-1)S^2}{\chi^2_{N-1}(1-\alpha/2)} < \sigma^2 < \frac{(N-1)S^2}{\chi^2_{N-1}(\alpha/2)}\right] = 1-\alpha$$

Thus, the $(1-\alpha)$ percent confidence interval for σ^2 is given by

$$\frac{(N-1)S^2}{\chi^2_{N-1}(1-\alpha/2)} < \sigma^2 < \frac{(N-1)S^2}{\chi^2_{N-1}(\alpha/2)}$$

Figure 11-2. The chi-square distribution partitioned so as to include $\alpha/2$ percent of the probability in each tail.



As an example of the use of this result, consider the random selection of 25 children from an elementary school who were given a complex concept-attainment task to learn. They were then given a standardized test for which the published norms stated that $\sigma = 20$. These 25 children had a sample standard deviation of 23. Using this information, we can determine the 95 percent confidence interval for σ^2 as follows:

1. $N - 1 = 25 - 1 = 24$
2. $S^2 = (23)^2 = 529$
3. $\chi^2_{N-1}\left(\frac{\alpha}{2}\right) = \chi^2_{24}(.025) = 12.40$
4. $\chi^2_{N-1}\left(1 - \frac{\alpha}{2}\right) = \chi^2_{24}(.975) = 39.36$
5. $\dfrac{(N-1)\,S^2}{\chi^2_{N-1}\left(1 - \frac{\alpha}{2}\right)} < \sigma^2 < \dfrac{(N-1)\,S^2}{\chi^2_{N-1}\left(\frac{\alpha}{2}\right)}$
6. $\dfrac{24(529)}{39.36} < \sigma^2 < \dfrac{24(529)}{12.40}$
7. $323 < \sigma^2 < 1024$

Since the determination of the exact confidence interval of σ is quite complex, it is generally sufficient to take the positive square roots of the end points of the confidence interval of σ^2 and use these to define the confidence interval for σ. Thus

$$17.98 < \sigma < 32.00$$

Because $\sigma = 20$ is included in this interval, it seems that the variability in scores is within expectation. Again, one should note that the probability that the *true* unknown variance is in the interval is either 1 or 0. The only thing that is known for sure is that if this procedure is used for all experiments or studies made, then 95 percent of the confidence intervals computed over this long range of experience will indeed contain the true unknown parameters.

11-13 INTRODUCTION TO EXPERIMENTAL DESIGN

Confidence intervals are not restricted to the placing of limits about parameters or differences in parameters of two populations; they can also be used to aid in the designing of an experiment or study. This other important use of confidence intervals is best understood by means of examples, and two are presented below.

Consider a hypothetical study of low SES (socioeconomic status) children in which one wishes to determine the average intelligence as measured by the Kuhlmann–Anderson intelligence test. Since testing involves students' classroom time

and the salary and time of professional testers it is desirable to estimate the unknown parameter values with as few students as possible. Since it is impossible to determine the value of μ exactly unless every child is tested, a tolerable amount of error in the estimate must be agreed upon. Suppose it is finally decided that the error should not exceed four IQ points. Translating this to a confidence-interval format, it means that the resulting confidence interval would have the following form:

$$\bar{X} - 4 < \mu < \bar{X} + 4$$

where $\bar{X}$ would be the mean of the sample of children tested. Writing the confidence interval out in its complete algebraic form, we have

$$\bar{X} - t_\nu\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{N}} < \mu < \bar{X} + t_\nu\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{N}}$$

and comparing the two equations, term by term, it is seen that

$$t_\nu\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{N}} = 4$$

This is a simple equation that involves the unknown sample size, the $\alpha/2$ percentile of the t distribution with ν degrees of freedom, and the value of S. If S and ν were known, then it would be possible to solve this equation for the unknown sample size.

Even though an estimated value that is not more than four points from the true value is wanted, it is necessary to accept a risk of a possible error by obtaining an interval that does not cover the true unknown value of the parameter. Suppose it is decided to control the chances of this kind of error to 5 percent. Examination of the t table shows that for moderately sized samples, $t_\nu(\alpha/2) = t_\nu(.025)$ and $t_\nu(.975)$ range in absolute value from about 2.20 to 2.02, and that for exceptionally large samples, $t = 1.96$. Since subjectivity dominates in planning, there is little harm in setting t equal to 2. Finally, one last bit of subjectivity must be considered—a number must be assigned to S. The standard deviation of the Kuhlmann–Anderson intelligence test is 15 points. Since these children are in a low SES school, it is reasonable to assume that the standard deviation for the school would be lower. As a compromise, let S be set equal to 10. Substituting these values into the equation, we can easily determine the sample size for the study. Thus

$$(2)\frac{10}{\sqrt{N}} = 4$$

$$20 = 4\sqrt{N}$$

$$\sqrt{N} = 5$$

$$N = 25$$

As part of this same hypothetical study, suppose there was interest in determining whether there was a difference in IQ scores between white and Negro children enrolled in the school. First, we make the same kinds of approximations that were made in the previous example; then we agree that the difference is unimportant unless the absolute average difference in IQ exceeds five points. The sample size for the study is easy to determine from the confidence interval for the difference in parameters for two independent populations, using the model for the difference between sample means. For equal-sized samples, the confidence interval of interest is given by

$$(\bar{X}_W - \bar{X}_N) - t_\nu\left(\frac{\alpha}{2}\right) S_p\sqrt{\frac{1}{N}+\frac{1}{N}} < \mu_W - \mu_N < (\bar{X}_W - \bar{X}_N) + t_\nu\left(\frac{\alpha}{2}\right) S_p\sqrt{\frac{1}{N}+\frac{1}{N}}$$

According to the conditions placed upon the interval by the researcher, the interval in this case should reduce to

$$(\bar{X}_W - \bar{X}_N) - 5 < \mu_W - \mu_N < (\bar{X}_W - \bar{X}_N) + 5$$

If one compares the two intervals, it follows that

$$5 = t_\nu\left(\frac{\alpha}{2}\right) S_p\sqrt{\frac{1}{N}+\frac{1}{N}}$$

Making the appropriate substitutions, we have

$$5 = (2)(10)\sqrt{\frac{2}{N}}$$

$$\frac{1}{4} = \sqrt{\frac{2}{N}}$$

$$\frac{1}{16} = \frac{2}{N}$$

$$N = 32$$

Thus, the minimum number of children necessary for the testing from each racial group is 32.

11-14 SUMMARY

In this chapter, confidence intervals, one of the most useful statistical models of modern statistical theory, were introduced. In some respects, the introduction of this statistical procedure to empirical research is the crowning achievement of mathematical statistics. The importance of these procedures was indirectly alluded to in the quotation from *Fortune* that introduced this chapter and in the examples that were used to illustrate the uses of confidence intervals.

In the typical experimental situation, a researcher has a parameter θ to be estimated. If only one universe is involved, the parameter may be μ, M, p, or σ^2. If two universes are involved, the parameter may be $\mu_1 - \mu_2$ or $p_1 - p_2$. Other parameters may also be of interest. Some of these other parameters will be seen in the following pages.

In general, the parameter θ is estimated from the sample statistics. In the one-sample case, these estimates are $\bar{X}$, $\hat{M}$, $\hat{p}$, and S^2. In the two-sample case, they are $\bar{X}_1 - \bar{X}_2$ and $\hat{p}_1 - \hat{p}_2$. If the variance of the sampling distribution of the estimators is known, then a $(1 - \alpha)$ percent confidence interval for θ is given by

$$\hat{\theta} - Z\left(\frac{\alpha}{2}\right)\sigma_{\hat{\theta}} < \theta < \hat{\theta} + Z\left(\frac{\alpha}{2}\right)\sigma_{\hat{\theta}}$$

If the variance of the sampling distribution of $\hat{\theta}$ is unknown, then the $(1 - \alpha)$ percent confidence interval for θ is given by

$$\hat{\theta} - t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\theta}} < \theta < \hat{\theta} + t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\theta}}$$

where ν = number of degrees of freedom used to estimate $\sigma^2_{\hat{\theta}}$, the unknown variance.

In the one-sample cases, $(1 - \alpha)$ percent confidence intervals for μ and p are given by

1. $\bar{X} - Z\left(\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{N}} < \mu < \bar{X} + Z\left(\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{N}}$

2. $\bar{X} - t_\nu\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{N}} < \mu < \bar{X} + t_\nu\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{N}}$

3. $\hat{p} - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}\hat{q}}{N}} < p < \hat{p} + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}\hat{q}}{N}}$

For all these intervals, random samples are required. For the confidence intervals for μ, one must assume that the parent population is normal or that N is sufficiently large to ensure that the sampling distribution of $\bar{X}$ is normal in form. For the confidence interval for p, one must assume that $Np > 5$ and $Nq > 5$ so that the binomial distribution can be adequately described by the normal distribution.

In the two-sample case, confidence intervals for $\mu_1 - \mu_2$ and $p_1 - p_2$ are given by

4. $(\bar{X}_1 - \bar{X}_2) - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$

5. $(\bar{X}_1 - \bar{X}_2) - t_\nu\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_p^2}{N_1} + \frac{S_p^2}{N_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_\nu\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_p^2}{N_1} + \frac{S_p^2}{N_2}}$

6. $(\bar{X}_1 - \bar{X}_2) - t^*_{\nu*}\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t^*_{\nu*}\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}$

and

7. $$(\hat{p}_1-\hat{p}_2)-Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1}+\frac{\hat{p}_2\hat{q}_2}{N_2}}<p_1-p_2<(\hat{p}_1-\hat{p}_2)+Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1}+\frac{\hat{p}_2\hat{q}_2}{N_2}}$$

For all four intervals independent random samples are required.

For the confidence intervals for $\mu_1-\mu_2$, it must be assumed that the parent populations are normal or that the sample sizes are large enough to ensure normality. When the variances are known, the interval (4) is appropriate. If the variances are equal and unknown, they can be estimated by

$$S_p^2=\frac{(N_1-1)S_1^2+(N_2-1)S_2^2}{(N_1-1)+(N_2-1)}$$

so that interval (5) is appropriate. In this case, $\nu=(N_1-1)+(N_2-1)$. If the variances are unequal and unknown, then the Welch–Aspin confidence interval (6) based on ν^* is acceptable. Finally, for the confidence interval (7) for p_1-p_2, one must know that $N_1p_1>5$, $N_1q_1>5$, $N_2p_2>5$, and $N_2q_2>5$ so that normality of distribution can be ensured. If any one of these confidence intervals extends across 0, then it is reasonable to conclude that $\mu_1=\mu_2$ or that $p_1=p_2$, depending upon which case is being studied.

An important confidence-interval procedure not covered in the previous cases involves repeated measurements on the sample elements. A frequent occurrence is in educational research where a pretest is given to a sample of subjects, a treatment is instituted, a posttest is given, and then difference scores are used to evaluate the method. In this case, paired observations are not independent, so that the above procedures are inappropriate. One way to counteract the lack of between-subject independence is to adopt a one-sample model of the difference scores $d_i=X_{1i}-X_{2i}$, $i=1,2,\ldots,N$. Since $\bar{d}=\bar{X}_1-\bar{X}_2$ and $\mu_d=\mu_1-\mu_2$, a confidence interval for μ_d is given by

8. $$\bar{d}-t_\nu\frac{S_d}{\sqrt{N}}<\mu_d<\bar{d}+t_\nu\frac{S_d}{\sqrt{N}}$$

where $\nu=(N-1)$ and $N=$ number of pairs. In this case one must assume that the observations come from normal universes, or that the distribution of the differences is normal, or that the sample size is sufficiently large so that the sampling distribution of $\bar{d}$ is normal.

Confidence intervals for σ^2 are based on the chi-square distribution. It was noted that if X is $N(\mu,\sigma^2)$, then a confidence interval for σ^2 is given by

9. $$\frac{(N-1)S^2}{\chi_\nu^2(1-\alpha/2)}<\sigma^2<\frac{(N-1)S^2}{\chi_\nu^2(\alpha/2)}$$

where $\nu=(N-1)$.

EXERCISES

***11-1.** Seventeen office workers were given a blindfold test in which they were asked to identify four different popular cola drinks. The numbers identified correctly were as follows: X: {1,2,2,0,4,0,1,2,1,1,4,0,1,0,0,1,1}

(*a*) Estimate the mean and variance of the universe.
(*b*) Set up 95 percent confidence intervals for μ and σ^2.
(*c*) How do the empirical values compare with the hypothetical values of Table 6-3?
(*d*) How do the office workers compare to the sophomores of Exercise 8-10?

***11-2.** In a study designed to test the effectiveness of special training using a talking typewriter, a group of seventh-grade remedial reading students were assigned to weekly training groups Their scores on a reading test following 20 weeks of training were as shown in the following table:

Number of sessions per week at typewriter		
NONE	ONE	FIVE
47	48	53
38	55	51
52	51	45
48	47	63
47	59	58
59	42	57
37	40	61
39		70
45		49

(*a*) Compute the means and standard deviations of each sample
(*b*) Set up 99 percent confidence intervals for $\mu_0 - \mu_1$, $\mu_0 - \mu_5$, and $\mu_1 - \mu_5$.
(*c*) Was the reading program effective? Why?

***11-3.** Set up the 95 percent confidence interval for $\theta = p_{c|m} - p_{c|f}$ of Exercise 4-9. On the basis of this evidence would you say that color blindness is independent of sex? Why?

***11-4.** Answer Exercise 7-1*c* in terms of the confidence-interval procedures of this chapter.

***11-5.** Set up the 95 percent confidence interval for the average amount of money spent on food for a family of four in Los Angeles. Use the statistics of Exercise 7-6.

***11-6.** Set up the 95 percent confidence interval for the differences in reinforcement conditions of Exercise 8-5. What does this tell you about the reinforcement conditions?

***11-7.** In a study of the influence of culture on family size, men in an Italian organization in New York City over age forty whose fathers were born in Italy were polled concerning the number of children they fathered and the number of children that their fathers fathered before age forty. The results were as shown in the table.

Number of children	*Frequency of club members*	*Frequency of fathers of club members*
0	5	0
1	12	3
2	36	9
3	10	29
4	7	16
5	2	12
6	0	1
7	0	0
8	1	4
9	2	1
Total	75	75

(*a*) Determine the mean and standard deviations for both sets of data.
(*b*) Set up the 95 percent confidence interval for differences in family size.
(*c*) What do these results suggest about the effects of culture on family size?
(*d*) One of the assumptions required for Exercise 11-7*b* is that the variables be statistically independent. Comment upon the assumption in this case. If the variables are not independent, what should one do?

***11-8.** In learning-studies, subjects are frequently used as their own controls in what is called a pretest-posttest design. The results of one such study in which form A of a mathematics reasoning test was used as a pretest and form B as a posttest were as shown in the table.

Student	*Pretest*	*Posttest*
Alissa	17	23
Bogard	23	37
Charlotte	19	22
Dorabella	16	18
Evalina	15	14
Fiora	17	30
Giorgetta	38	32
Harold	17	18
Ina	22	25
Jolanda	25	29
Kirk	37	30

(*a*) Set up a 95 percent confidence interval for the mean improvement.
(*b*) Was there an improvement?
(*c*) Set up a 95 percent confidence interval for σ_d^2.
(*d*) What assumptions have you made? Are they reasonable?

***11-9.** Often, one consequence of teaching is the decrease in variability among students following special training. With this in mind, consider the following statistics, which were generated in a concept-learning experiment. The control subjects were told to learn the task in as few trials as possible. The experimental subjects were given cues on how to learn the task. The dependent variable of the study is the number of trials it took the subject to make three correct consecutive responses.

Group	*Sample size*	*Average*	*Standard deviation*
Control	18	19.7	8 3
Experimental	23	14.6	2 7

(*a*) Set up a 95 percent confidence interval for the differences in mean number of trials.
(*b*) What was the effect of the training?
(*c*) Set up 95 percent confidence intervals for σ_c^2 and σ_e^2.
(*d*) Would you say that the variances of the universes are equal or unequal? Why?
(*e*) What assumptions have you made in the analysis of these data?

***11-10.** In a study of the effectiveness of a teaching program in foreign languages, 18 first-year students majoring in German were given a difficult passage of scientific German to read. The time taken to read the material was determined for each student. The sample statistics were as follows: $\bar{X} = 228$ seconds and $S = 56$ seconds.

(*a*) Set up the 95 percent confidence interval for μ.
(*b*) How do these results compare to the results for second-year students discussed in Section 11-4?
(*c*) Set up 95 percent confidence intervals for σ^2 for the first- and second-year students.
(*d*) What can you say about the variability in scores for these two sets of students?
(*e*) What assumptions have you made?

12 HYPOTHESIS TESTING IN THE ONE-SAMPLE MODEL

As the 1968 presidential campaign moves into the home stretch, the public opinion polls still favor Richard Nixon over Hubert Humphrey, but there is increasing uncertainty about it all.

The two most widely published reports, Harris and Gallup, both show Nixon ahead and both show Humphrey gaining.

Both Harris and Gallup polls have a built-in possibility of statistical error of about 3 to 4%. Unknown factors might even increase the possibility of error.

12-1 INTRODUCTION TO THE GENERAL THEORY OF HYPOTHESIS TESTING

Many of the research questions encountered in behavioral research can be solved quite adequately by the use of confidence intervals. While this may not be obvious or readily apparent, reasons for stating this will become more apparent in the following pages. As will be seen, confidence intervals, in addition to providing interval estimates of unknown parameters, can also be used to test hypotheses about parameters. For the most part, behavioral scientists have not utilized this property of confidence intervals to any great extent, since many were trained at a time when greater allegiance was given to hypothesis testing and the rejection of the "null hypothesis." Modern research methods emphasize confidence-interval procedures instead of hypothesis-testing models In any case, the confidence-interval approach is better understood and appreciated if the theory of hypothesis testing is understood.

To begin the discussion of hypothesis testing, consider the following hypothetical example. A group of college sophomores in a large psychology class were given an achievement test on personality theory. The scores for the students were transferred to IBM cards and processed on a computer. The resulting statistics were as reported in Table 12-1. The corresponding normal distributions with parameters equal to these sample values are shown graphically in Figure 12-1.

After the data were processed, the cards were placed in a cardboard box and stored. Two weeks later, further data analysis was required. When the cards were being removed from their box, they were accidentally dropped on the floor and were thoroughly mixed. Unfortunately, the research assistant who had done the original card punching failed to punch a code for the very important variable sex. Thus, it was impossible to determine the sex of the student associated with each card. Examination of the probability distribution suggests that all cards with scores below 70 are associated with females and that all cards with scores above 110 are associated with males. For scores between 70 and 110 it is difficult to say which sex should be associated with each card. Low-valued scores in this middle range are generally associated with females while high-valued scores in this same range are usually associated with males. Intuition suggests that it should be possible to determine a hypothetical score that could be used as a cutoff point to simplify the

Figure 12-1. Theoretical distribution of scores made by males and females on a hypothetical personality test given to college sophomores

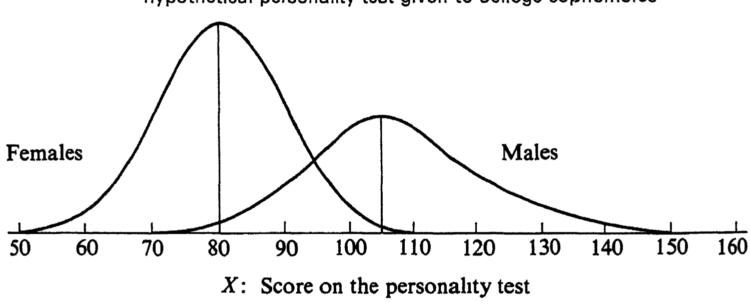


Table 12-1. Sample statistics on a hypothetical personality test given to college sophomores.

Sex of student	*Average score*	*Standard deviation of score*	*Number of students*
Male	108	14	220
Female	81	9	197

classification of each card as representing a female or male score. For example, it might be said that all scores of 79 or less are female scores and that all scores of 80 or more are male scores. Without doubt this operational definition of sex will produce errors in classification, since some female cards will be incorrectly identified as male cards and some of the male cards will be incorrectly identified as female cards. Intuition suggests that there is some one best rule that will minimize these two kinds of errors. An attempt to identify this rule might proceed as follows.

Suppose it is decided to use the rule mentioned. If the scores have a normal distribution, then

$$P_F = P[\text{misclassifying a female}]$$
$$= P[X > 79 \,|\, \text{subject is female}]$$

Making a correction for continuity, we have

$$P_F = P[X > 79.5 \,|\, \text{subject is female}]$$
$$= P\left[Z > \frac{X - \mu_F}{\sigma_F} \,\middle|\, X = 79.5,\ \mu_F = 81,\ \sigma_F = 9\right]$$
$$= P\left[Z > \frac{79.5 - 81}{9}\right]$$
$$= P\left[Z > \frac{-1.5}{9}\right]$$
$$= P[Z > -.17]$$
$$= .567$$

In a like manner,

$$P_M = P[\text{misclassifying a male}]$$
$$= P[X \leqslant 79 \,|\, \text{subject is male}]$$

Making the correction for continuity gives

$$P_M = P[X < 79.5 \mid \text{subject is male}]$$

$$= P\left[Z < \frac{X - \mu_M}{\sigma_M} \,\middle|\, X = 79.5,\ \mu_M = 108,\ \sigma_M = 14\right]$$

$$= P\left[Z < \frac{79.5 - 108}{14}\right]$$

$$= P\left[Z < \frac{-28.5}{14}\right]$$

$$= P[Z < -2.05]$$

$$= .020$$

Thus, the total probability of error is given by

$$p = \frac{N_F P_F + N_M P_M}{N_F + N_M}$$

$$= \frac{197(.567) + 220(.020)}{197 + 220}$$

$$= \frac{112 + 4}{417}$$

$$= .28$$

The largest part of the probability of error results from the large misclassification of females. Fortunately it is possible to reduce the error of misclassifying a female by moving the cutoff point to a higher value. Thus, by trial and error it is possible to find an optimum rule for classifying these students. In this case the optimum rule is obtained when the cutoff point is about 94.5. For this cutoff point,

$$P_F = P[\text{misclassifying a female}]$$

$$= P[X > 94.5 \mid \text{subject is female}]$$

$$= P\left[Z > \frac{94.5 - 81}{9}\right] = P[Z > 1.50]$$

$$= .067$$

and

$$\begin{aligned}P_M &= P[\text{misclassifying a male}]\\&= P[X < 94.5|\text{subject is male}]\\&= P\left[Z < \frac{94.5 - 108}{14}\right] = P[Z < -.96]\\&= .169\end{aligned}$$

so that the total probability of misclassification in this case is given by

$$\begin{aligned}p &= \frac{197(.067) + 220(.169)}{197 + 220}\\&= \frac{13 + 37}{417}\\&= .12\end{aligned}$$

Although the probabilities of error have been switched around, collectively they are smaller than formerly. Thus, the probability of making false statements has been reduced.

Even though this example represents a simplified version of statistical classification theory, it also illustrates some of the basic elements of the statistical theory of hypothesis testing. In this theory, one has a hypothesis that is to be tested and compared to an alternative hypothesis. In this example, the hypothesis tested is that the individual associated with the card being examined is a male. The alternative hypothesis is that the individual is not a male. The hypothesis to be tested is generally denoted in the literature by H_0, the alternative hypothesis being denoted by H_1. In this example, these would be symbolized in the following manner. X has been randomly selected from a universe for which H_0 is true or for which H_1 is true, where H_0 and H_1 are specified by

$$\begin{aligned}&H_0: \mu_0 = 108 \text{ and } \sigma_0^2 = 14^2\\&H_1: \mu_1 = 81 \text{ and } \sigma_1^2 = 9^2\end{aligned}$$

Thus, it is seen that one is dealing with a dichotomous decision theory in that a choice is being made between two different statistical hypotheses or interpretations for the data. For this sort of decision process, one always entails the risk of saying that H_0 is true when really H_1 is true, or vice versa. A good statistical test minimizes the chances of making these errors; as yet, there is no test that completely eliminates them. As a result, they must always be reckoned with.

As indicated by the previous example, a hypothesis is a guess or statement about nature that needs verification. A *statistical hypothesis* is a comparable statement about a probability distribution. Usually the statements are made in terms of

expected values or variances, though other population measures may be used. A statistical hypothesis is subjected to a *statistical test*, which is, in essence, a rule that states when a statistical hypothesis should be rejected or not rejected. When a true statistical hypothesis is incorrectly rejected, an error has occurred. This error is called a *type I error*. If, on the other hand, a false hypothesis is not rejected, a second kind of error has occurred. This error is called a *type II error*. A good statistical rule minimizes the probabilities of making these errors. These probabilities are denoted in the literature as follows:

$$P[\text{type I error}] = \alpha$$
$$P[\text{type II error}] = \beta$$

These statements and associated probabilities are summarized in Tables 12-2 and 12-3. The probability $(1 - \beta)$ is called the power of the test. It is the probability of rejecting H_0 when H_0 is false. It is synonymous with the biologist's use of the word *power* in rating a laboratory microscope A microscope with high power enables the biologist to enlarge small objects so that they may be studied in detail. A powerful statistical test enables a researcher to see small population differences more clearly so that rejection of a false hypothesis is accentuated and made with greater confidence.

Table 12-2. Decisions and errors in hypothesis testing.

Decision concerning H_0	*Hypothesis under test*	
	H_0 IS TRUE AND H_1 IS FALSE	H_0 IS FALSE AND H_1 IS TRUE
Reject H_0	Type I error	Correct decision
Do not reject H_0	Correct decision	Type II error

Table 12-3. Probabilities of decisions in hypothesis testing

Decision concerning H_0	*Hypothesis under test*	
	H_0 IS TRUE AND H_1 IS FALSE	H_0 IS FALSE AND H_1 IS TRUE
Reject H_0	α	$1 - \beta$
Do not reject H_0	$1 - \alpha$	β

To further illustrate the concepts, consider the following hypothetical example. Suppose that a school is located in a high SES area and it is believed that in the school the mean achievement score on a standardized test will be 120 instead of 90, as it is for the general population. Let it also be assumed that the standard deviation of scores is equal to 10 regardless of what value the average assumes. In a hypothesis-testing model, it is customary to adopt a negative view of the situation and to consider the hypothesis that states that the average score is not 120 but 90. Under this model the alternative hypothesis is that $\mu = 120$. Notice that the rejection of H_0 establishes the truth of H_1. The major reason for adopting this inverted procedure for establishing the truth of H_1 is that one can disprove statements with ease simply by finding contrary evidence. The proof of a proposition based on empirical data is logically unattainable, since it is generally impossible to gather all information and evidence concerning the truth or falsity of a statement. Since one contradictory bit of evidence concerning the truth or falsity of a testable hypothesis is sufficient to declare a statement as being false, one can establish the truth of a statement by denying or rejecting its alternative hypothesis. Thus, if one can reject H_0, then it follows that H_1 has been established.

For the purposes of exposition, assume that it has been decided to test only *one* student at the school and on the basis of the observed score to make a decision as to the truthfulness of H_0 or H_1. If X represents the score on the test, then the statistical hypothesis of interest and its alternative are given by

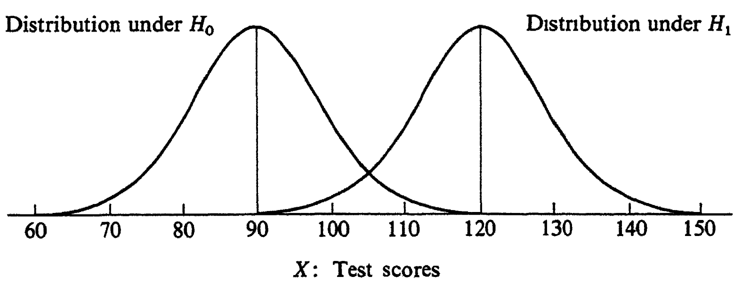
$$H_0: \mu_0 = 90$$
$$H_1: \mu_1 = 120$$

where it is assumed that X has a normal distribution with $\sigma = 10$. The graphic representation of the statistical model is shown in Figure 12-2.

In order to test the hypothesis, one must have a decision rule that will lead to the right conclusion. Suppose the following decision rule is chosen

$$\text{D.R.: reject } H_0 \text{ if } X > 120$$

Figure 12-2. Hypothetical distributions of test scores



This rule produces a test such that the probability of a type I error is very close to 0. The probability of a type II error for this test is represented by the area that falls below 120 in the probability distribution defined by H_1. In this case the probability of a type II error is equal to .5 Because there is a large probability of not rejecting a false statement, it must be concluded that the rule is not very good and therefore a new rule should be tried.

As a new rule, consider

D.R.: reject H_0 if $X > 110$

In this case,

$$\alpha = P(\text{type I error}) = .0228$$
$$\beta = P(\text{type II error}) = .1587$$

This rule is better than the first rule because the sum of the probabilities of the errors is smaller ($.1815 < .5$). However, the probability of being in error is still large, therefore calling for another rule.

As a possible rule, consider

D.R.: reject H_0 if $X > 100$

Here one has

$$\alpha = P(\text{type I error}) = .1587$$
$$\beta = P(\text{type II error}) = .0228$$

There is not much difference between the last two rules, except that the probability of a type I error is greater in the second case and the probability of a type II error is smaller. Therefore, consider the compromise rule

D.R.: reject H_0 if $X > 105$

With this rule, the probability of a type I error is given by $P[X > 105 | H_0 \text{ is true}] = P[Z > 1.5 | H_0] = .0668$. At the same time, the probability of a type II error is given by $P[X < 105 | H_1 \text{ is true}] = P[Z < -1.5 | H_1] = .0668$. Thus, the total probability of error is .1336, which is smaller than that produced by any of the other rules.

Suppose a student is given the test and he obtains a score of 114. One would now conclude that H_0 is false and that H_1 is true The probability of having made a correct decision is 1 or 0 since either a correct decision was made or it was not.

Notice that as the decision rule is changed the value of α, the probability of a type I error, changes. Also note that α can be made as small as desired at the expense of making β, the probability of a type II error, very large. In like manner one can make β very small and α very large. Unfortunately, with a small sample size it is impossible to reduce both errors simultaneously, since as one decreases, the other increases. Because of this, it is customary to compromise and proceed in much the same manner as one does in choosing the confidence coefficient for a confidence

interval. Operationally, this means that the probability of one of the errors is controlled. In this control a decision is made as to the magnitude of a type I error one is willing to tolerate. Most researchers, depending on the situation, choose either .05 or 01. Suppose, in the example, that it is decided to set α equal to .05. The decision rule for this α is given by

$$\text{D.R.: reject } H_0 \text{ if } X > \mu + 1.645\sigma = 90 + 1.645(10) = 106.45$$

which is 1.645 standard deviations above the mean. With this decision rule the probability for a type II error is given by

$$\begin{aligned}\beta &= P[X < 106.45 | H_1] \\ &= P\left[Z < \frac{106.45 - 120}{10}\right] \\ &= P[Z < -1.355] \\ &= .0894\end{aligned}$$

As a result, the power of the test is given by

$$\begin{aligned}P &= 1 - \beta \\ &= 1 - .0894 \\ &= .9106\end{aligned}$$

If one wanted to reduce β and hold α constant, then it would be necessary to increase the sample size. To see the effects of increased sample size on β, suppose that N were to be increased to 2. With more than one observation a new criterion is needed. The sample average is a good prospect, since it is known that $\bar{X} = \frac{1}{2}(X_1 + X_2)$ is the best estimate of the unknown parameter. As a result, the decision rule would have to be stated in terms of $\bar{X}$. To evaluate this decision rule one could not use the distribution of X. Instead it would be necessary to use the sampling distribution of the averages when H_0 and H_1 are true. For H_0 the expected value of the means will be given by $\mu_0 = 90$, while the standard deviation of the sampling distribution of the means is given by

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{10}{\sqrt{2}} = 7.07$$

The expected value of the means if H_1 were true would be given by $\mu_1 = 120$, while the standard deviation of this distribution of means would be exactly the same as the other distribution. The two sampling distributions of $\bar{X}$ would appear as shown in Figure 12-3. Notice that in Figure 12-3 much of the overlap is gone. This reduction has been accomplished by increasing the sample size to 2. In essence, increasing the sample size has separated the sampling distributions of the criterion variable

under H_0 and H_1. This means that the test has become more powerful. If, again, $\alpha = .05$, the rejection rule is given by

$$\begin{aligned} \text{D.R.}\quad \text{reject } H_0 \text{ if } \bar{X} &> \mu_0 + 1.645\sigma_{\bar{X}} \\ &= 90 + 1.645(7.07) \\ &= 101.63 \end{aligned}$$

The probability of concluding that H_0 is true when really H_1 is true is equal to the probability that $\bar{X} < 101\ 63$ under the condition that H_1 is true. This is given by

$$\begin{aligned} \beta &= P[\bar{X} < 101\ 63 | H_1] \\ &= P\left[Z < \frac{101.63 - 120}{7.07}\right] \\ &= P[Z < -2.60] \\ &= .0047 \end{aligned}$$

Thus, the probability of a type II error is quite small. The power of the test is given by

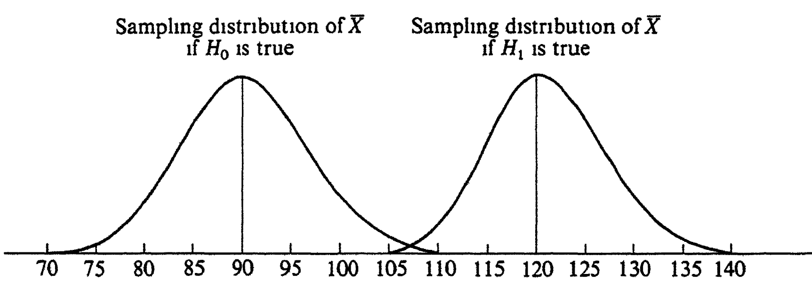
$$\begin{aligned} P &= 1 - \beta \\ &= 1 - .0047 \\ &= .9953 \end{aligned}$$

In this case, one is almost certain not to make a type II error.

The discussion up to this point has been restricted to the testing of a *simple hypothesis* against a simple alternative. This is usually denoted by

$$H_0: \mu = \mu_0$$

Figure 12-3. Hypothetical distributions of means for samples of size 2.



against the alternative

$$H_1: \mu = \mu_1$$

Suppose that $\mu = 90$ had been false, so that a false hypothesis had been tested. Furthermore, suppose that the true parameter was really given by $\mu = 80$ but that this fact was unknown to the researcher. For sample sizes of 2, the decision rule was given by

D.R.: reject H_0 if $\bar{X} > 101.14$

If this were indeed the case, then α associated with the true situation would be smaller than .05 and β would have remained unaffected. Thus, even though the wrong hypothesis may have been tested, the probability of a type I error is even lower. Furthermore, it would be for any value of $\mu < 90$. This means that it might be wise to test the *composite hypothesis* H_0 $\mu \leqslant 90$ against the simple alternative $H_1: \mu = 120$. In a general case this would mean that H_0 would be replaced by

$$H_0: \mu \leqslant \mu_0$$

which would then be tested against

$$H_1: \mu = \mu_1$$

$\mu = \mu_0$ is called the least favorable outcome for μ and α represents the maximum protection that is provided for this least favorable value of μ.

In like manner, the least favorable value for μ under the alternative hypothesis might be 120, while in reality μ may exceed 120. If one plans minimum β protection against this least favorable alternative, then the protection against alternatives exceeding it is greater. When this is done, a composite hypothesis can be tested against a composite alternative. This would be denoted as

$$H_0: \mu \leqslant \mu_0$$
$$H_1: \mu \geqslant \mu_1$$

Generally, the value of μ_1 is unknown and so the alternative cannot be stated so specifically. Because of this, the typical hypothesis-testing situation is given by

$$H_0: \mu \leqslant \mu_0$$

versus the alternative

$$H_1: \mu > \mu_0$$

or

$$H_0: \mu \geqslant \mu_0$$

against the alternative

$$H_1: \mu < \mu_0$$

Finally, it should be noted that the hypothesis tested is stated before the experiment is begun, even before the experimenter goes near an experimental subject.

According to contemporary usage, $\bar{X}$ in the previous example is called a *test statistic*, since numerical values are compared to the set of rejection values defined by the decision rule. In the literature, $\bar{X}$ is not used as a test statistic; instead, its transformed Z value is preferred. In the example, the final decision rule for $N=2$ was given by rejecting H_0 if $\bar{X} > 101.63$. If the test statistic $\bar{X}$ is replaced by the test statistic $Z = (\bar{X} - 90)/7.07$, then the corresponding decision rule is given by

D.R.: reject H_0 if $Z > 1.645$

For the general hypothesis-testing model,

$$H_0: \mu \geqslant \mu_0$$

against the alternative

$$H_1: \mu < \mu_0$$

the test statistic for a sample size N is given by

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{N}}$$

and the α decision rule is given by

D.R.: reject H_0 if $Z > Z(1 - \alpha)$

Of course, if the directions of the inequalities were changed, then the decision rule would be

D.R.: reject H_0 if $Z < Z(\alpha)$

12-2 ONE-SAMPLE TEST FOR EXPECTED VALUES WHEN THE POPULATION VARIANCE IS KNOWN

As part of a hypothetical curriculum study in science education it was decided to use Penfield's Electrical Science Test (PEST) to evaluate the efficacy of a new method for the teaching of electrical circuitry. Published norms for the test stated that 80 is the average score and that the standard deviation of the test is 12. Students were trained with a new teaching method and at the end of the training period 29 children were given the test. If the new teaching method were no better than the old method, and if the test subjects were similar to those on which the test was normed, it would be reasonable to assume that the sample distribution of scores would have a mean value close to 80, but if the method were effective, the mean score of the sample would exceed 80. The latter possibility could be considered as evidence supporting the belief that the method, when applied to other groups of children selected from similar populations, could also produce scores that exceed 80 on the average.

Reformulating these statements in a hypothesis-testing format, one has

H_0: The new method of teaching electrical circuitry has no effect upon the scores obtained on the Penfield Electrical Science Test

versus

H_1: The new method of teaching produces elevated scores on the Penfield Electrical Science Test

Restating these two hypotheses in a statistical form, one has

$H_0: \mu_0 = 80$

versus

$H_1: \mu_1 > 80$

Notice that the alternative hypothesis is a composite hypothesis, since the exact value of μ_1 is not specified by any numerical value. This, unfortunately, is a very common situation in educational research since the exact effects of any new teaching method cannot be stated with any degree of certainty. One of the goals in educational research is to estimate this effect.

When the Penfield test was created, one of the requirements placed upon its final form was that the distribution of test scores be bell shaped and symmetrical In this instance the manufacturers were somewhat successful in achieving this goal. As a result, there is good reason to believe that the theoretical distribution of means of all possible samples of size 29 should be quite close to normal because the underlying variable is close to normal in form. Since the performance of each child on the test was independent of the performance of the other children, a simple test statistic of the form

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{N}}$$

could be used to test the hypothesis of interest. Since the consequences of a type I error were not exceptionally great, it was decided to control the risk of a type I error at .05. This means that H_0 would be rejected if the value of the test statistic exceeds 1.645. In this particular case, the average score of the 29 children was equal to 85.3, so that

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{N}} = \frac{85.3 - 80}{12/\sqrt{29}} = \frac{5.3}{2.2283} = 2.38$$

Since this value exceeds 1.645, H_0 is rejected; therefore there is definite reason to believe that training does indeed raise the average score of children who are taught in this manner, or that these children are more able than was the group on which the test was standardized. In any case, it should be noted that the value of μ is still

unknown. The best point estimate of this unknown value is given by $\bar{X} = 85.3$, which, of course, is the unbiased and efficient estimate of μ. Certainly this estimate can be used in future planning programs; however, it is very informative to know what limits should be placed on the unknown population parameter value Since the rejection region of this test of hypothesis is defined by the upper tail of the $N(0,1)$ distribution, a one-tailed confidence interval is appropriate. In this case, the one-tailed confidence interval is given by

$$\begin{aligned} \mu &> \bar{X} - Z(1-\alpha)\,\sigma_{\bar{X}} \\ &> 85.3 - 1.645(2.2283) \\ &> 85.3 - 3.7 \\ &> 81.6 \end{aligned}$$

Thus, on the average, one can feel fairly certain that this particular teaching method tends to increase the average score by at least 1.6 points or about one-eighth of a standard deviation. At this point the science curriculum expert must decide whether such an increase is important. Just because it is statistically significant does not mean that it is meaningful. The meaningfulness of the increase is not a statistical question but a curriculum question, which the curriculum expert must react to.

In this particular case, it was decided that an average increase of 5.3 was important but that an average increase of 6 points would represent a very important increase in average score. This immediately raises the question as to whether this particular study would have identified this increase with any reasonable level of power. This is easy to determine. In this case,

$$\begin{aligned} \text{Power} &= P(\text{rejecting } H_0 \text{ if } H_1 \text{ is true}) \\ &= P(Z > 1.645 \mid \mu_1 = 86) \\ &= P(\bar{X} > \mu_0 + 1.645\sigma_{\bar{X}} \mid \mu_1 = 86) \\ &= P(\bar{X} > 80 + 1.645(2.2283) \mid \mu_1 = 86) \\ &= P(\bar{X} > 83.7 \mid \mu_1 = 86) \\ &= P\left(Z > \frac{83.7 - 86}{2.2283}\right) \\ &= P\left(Z > \frac{-2.3}{2.2283}\right) \\ &= P(Z > -1.03) \\ &= .85 \end{aligned}$$

This is a rather high probability for rejecting a false hypothesis. This further supports the subjective feeling that the observed difference is indicative of a real difference attributable to the new teaching method. In fact, one might seriously consider an

adoption of this new method in the classes of the school district in which electrical circuitry principles are taught.

12-3 ONE-SAMPLE TEST FOR EXPECTED VALUES WHEN THE POPULATION VARIANCE IS UNKNOWN

In the study described in Section 12-2, it was assumed that the variability of scores for the new method was identical to the variability of the normed standardized test. For many educational studies, this assumption is difficult to justify since some educational methods tend to spread students along the scale of the criterion variable while other methods tend to produce homogeneity among the tested subjects. When this occurs, one should not use the statistical test of the previous section. Instead, one should employ the information available in the sample concerning the variability in scores for the new method.

For this model, the hypothesis to be tested and its alternative are still given by

$$H_0\cdot \mu_0 = 80$$

versus the alternative

$$H_1: \mu_0 > 80$$

Because the variance of the population is unknown, a test based upon the statistic $Z = (\bar{X} - \mu_0)/(\sigma/\sqrt{N})$, which is normally distributed with mean of 0 and variance of 1 when H_0 is true, is not possible. However, as noted earlier,

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{N}}$$

has a t distribution with $\nu = N - 1$ degrees of freedom when the hypothesis tested is true. In this case, $\nu = 29 - 1 = 28$, so that with $\alpha = .05$, the decision rule for the test is given by

D.R.: reject H_0 if $t > 1.70$

Suppose that the standard deviation for the 29 test scores was given by $S = 10.3$; then

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{N}}$$

$$= \frac{85.3 - 80}{10.3/\sqrt{29}} = \frac{5.3}{1.9126}$$

$$= 2.77$$

Since $t = 2.77$ is in the rejection region, there is reason to doubt the hypothesis that $\mu = 80$ Thus, the same decision is made as before. The one-sided confidence interval for the unknown parameter value is given by

$$\mu > \bar{X} - t_\nu(1 - \alpha)\mathrm{SE}_{\bar{X}}$$

$$> 85.3 - 1.70(1.9126)$$

$$> 85.3 - 3.3$$

$$> 82.0$$

This example illustrates the importance of variance upon hypothesis testing in that one should pay close attention to the appropriate model. In this case, this importance is expressed in the question "Should one assume σ is known or unknown?" If one adopts a conservative point of view, then it is reasonable to assume that σ is unknown. This means that one would normally prefer the t model over the Z model.

This example also illustrates a common problem of educational research. There are many reasons, other than the effectiveness of the training program, why the tested hypothesis has been rejected. Some competing explanations are that the students in the study are better prepared for the task than were the students on which the test was normed, or that the teacher and students may have been aware of the purposes of the study and thereby paid particular attention to improvement in learning. This latter explanation is often referred to as the Hawthorne effect, and is generally a possible contender in the explanation of rejected hypotheses.

One of the assumptions that must be made when performing the t test is that the parent population is normal in form. Fortunately, this assumption is not too important. As predicted by the central limit theorem, the sampling distribution of $\bar{X}$ tends to a normal form even if X, the underlying variable, is not normal. As a result, it is said that the t test is robust with respect to the lack of normality. This means that the actual probability of a type I error is very close to the nominal or specified probability of a type I error, with the closeness increasing as the sample size increases. Thus, for large sample size, the normality assumption is almost unnecessary. For small samples, the assumption takes on greater importance, especially if the parent population is not symmetrical, since the approximation of the sampling distribution of $\bar{X}$ to the normal may not be adequate.

There is also some evidence that the power of the test and the probability of a type II error are not significantly affected by the lack of normality, provided that the sample is large. The case for small samples is not too clear. Operationally, this means that rejection of H_0 with the t test is a strong statement, while nonrejection may reflect poor power resulting from the lack of normality and not necessarily mean that H_0 is true.

12-4 ONE-SAMPLE TEST FOR THE PARAMETER OF A BINOMIALLY DISTRIBUTED VARIABLE

In the survey on de facto segregation of schools described earlier, a random sample of adults was asked "It has been suggested that school boundaries be changed so that the percentage of nonwhite and white children in these schools would be more like the percentage for the entire school system. 1. I agree_____ 2. I disagree_____ 3. I am not sure_____."

Of the 733 respondents, 467 agreed to the boundary change integrating the schools. As a result, one would like to know whether this reflects a significant positive attitude toward school integration within the entire community, since $\hat{p} = \frac{467}{733} = .637$ of the respondents looked with favor upon the proposition. If the "I am not sure" responses are combined with the "I disagree" responses, then each person's response represents a sampling from a Bernoulli distribution $B(1,p)$, so that the sampling distribution of $T = X_1 + X_2 + \cdots + X_{733}$ is characterized by $B(733, p)$ where $p = P(\text{agree})$. If there is a positive attitude in the community, then $p > \frac{1}{2}$ If there is a negative attitude, then $p < \frac{1}{2}$. This suggests that the set of alternative values for p that are of interest are contained in two mutually exclusive subsets of the interval $0 \leqslant p \leqslant 1$. When this occurs, it is said that the alternative hypothesis defines a two-tailed test of hypothesis. In this case, a two-tailed test of hypothesis is appropriate even though the observed $\hat{p}$ exceeds $\frac{1}{2}$, because the hypothesis of indifference among the population members is made *prior* to the collection of data. The hypothesis is not stated after data are collected since data are never permitted to suggest hypotheses. Finally, if responses are determined independently by each member of the sample, then the total number of "I agree" responses is binomial. Therefore, the following hypothesis-testing model is appropriate:

H_0: The community is equally divided on the changing of school boundaries to achieve racially balanced schools

H_1: The community is not equally divided on this issue

In statistical terms, the hypothesis tested and its alternative are

$$H_0: p = p_0, \text{ where } p_0 = \tfrac{1}{2}$$

versus the alternative

$$H_1: p \neq \tfrac{1}{2}$$

If $Np_0 > 5$ and $Nq_0 > 5$, this can be tested by either of the following two test statistics:

$$Z = \frac{T - Np_0}{\sqrt{Np_0 q_0}} = \frac{T - 733(\frac{1}{2})}{\sqrt{733(\frac{1}{2})(\frac{1}{2})}}$$

or

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / N}} = \frac{\hat{p} - \frac{1}{2}}{\sqrt{(\frac{1}{2})(\frac{1}{2})/733}}$$

with the decision rule

D.R.: reject H_0 if $Z < -1.96$ or if $Z > 1.96$

If $Np_0 < 5$ and $Nq_0 < 5$, then the exact testing procedure of Section 5-5 must be used. For intermediate size samples, the continuity correction is suggested. In this case, the sample size is large enough so that the continuity correction can be ignored. Thus

$$Z = \frac{467 - 733(\frac{1}{2})}{\sqrt{733(\frac{1}{2})(\frac{1}{2})}} = 7.42$$

or

$$Z = \frac{.637 - \frac{1}{2}}{\sqrt{(\frac{1}{2})(\frac{1}{2})/733}} = 7.42$$

Since $Z = 7.42$ is larger than $Z = 1.96$, the hypothesis is rejected. Thus, it is concluded that the community is not equally divided on this issue.

The best point estimate of the unknown p value is .637 and the 95 percent confidence-interval estimate is given by

$$p = \hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{N}} = .637 \pm 1.96\sqrt{\frac{.637(.363)}{733}}$$

$$= .637 \pm .035$$

which can be written as $.602 \leqslant p \leqslant .672$. Although this confidence interval is similar in form to the one that would be obtained if $p = \hat{p} \pm 1.96\sqrt{p_0 q_0 / N}$ were used, the use of $SE_{\hat{p}} = \sqrt{\hat{p}\hat{q}/N}$ in place of $\sigma_{\hat{p}} = \sqrt{p_0 q_0 / N}$ is mandatory because the hypothesis H_0: $p = p_0 = \frac{1}{2}$ has been rejected. If the assertion $p = p_0$ has been denied, then it is no longer correct to use $p_0 q_0 / N$ as the variance of $\hat{p}$. Instead, the variance of $\hat{p}$ is estimated by replacing (p,q) in $\text{Var}(\hat{p}) = pq/N$ with the point estimates $\hat{p}$ and $\hat{q}$. Thus, the estimated variance of p is given by the squared standard error $SE_{\hat{p}}^2 = \hat{p}\hat{q}/N$.

As a result of this test and the limits of the corresponding confidence interval, there is very good reason to believe that more than 60 percent of the population favors the changing of school boundaries to achieve school integration.

In this example, the correction for continuity was not made even though the underlying variable was discrete. If it had been made, Z would have remained large and H_0 would have been rejected. However, if Z had been close to ± 1.96, then the correction for continuity should not have been ignored since acceptance or rejection of H_0 will be dependent upon the best approximation to the true unknown binomial probability For this reason, it is suggested that one always use the correction for continuity or, if this is not done, that one use it at least for those Z values close to $-Z(\alpha/2)$ or $+Z(\alpha/2)$.

Notice that the difference between the hypothesis value and the observed percentage is 7.42 standard deviation units. Older textbooks say that this represents a highly significant difference, since $P(Z > 7.42) < .0001$. This sort of interpretation, still made by some contemporary research workers, illustrates the fuzzy kind of thinking that still permeates present research The person who makes such a statement is assuming that the hypothesis under test has a probability of being a correct statement. As has been indicated repeatedly, a statement can only be true or false and the probabilities of these two events are either 1 or 0. When one does a statistical test, one is using a procedure that helps in making the decision as to whether one should behave as though the hypothesis is true or false. He does not establish the truth or falsity of the statement by rejecting or failing to reject the hypothesis. Regardless of the decision that is made by the researcher, the final truth of the statement must and will remain unknown. Only a complete census of the population can establish its truth or falsity.

The fact that $Z = 7\ 42$ says something else about the study. In essence, it says that the power for rejection is extremely high and perhaps very close to unity. In survey research, large samples frequently provide more than minimal power to execute the test. Furthermore, information is frequently obtained on many variables, not all of which are dichotomous. The large number of informants for this survey was demanded for other reasons and the large power for rejection is quite incidental. Even so, it is informative to consider what size sample would have sufficed for the rejection of the hypothesis $p = \frac{1}{2}$ versus the alternative $p > \frac{6}{10}$, with a probability of a type I and type II error both controlled at .05.

Since power is related to the true value of p, it suffices to determine the power for the least significant difference of interest. In this case, it occurs when $p = \frac{6}{10}$. For any value of $p > \frac{6}{10}$, the power will be greater for any fixed sample size. With the conditions ($p_0 = \frac{1}{2}$, $\alpha = .05$) and ($p_1 = \frac{6}{10}$, $\beta = .05$), and with 1.96 replaced by 2 to simplify the arithmetic,

$$\begin{aligned}
\text{Power} &= P(\text{rejecting } H_0 \text{ when } H_1 \text{ is true}) \\
&= P(Z < -2 \cup Z > +2 \mid p_1 = \tfrac{6}{10}) \\
&= P(Z < -2 \mid p_1 = \tfrac{6}{10}) + P(Z > 2 \mid p_1 = \tfrac{6}{10})
\end{aligned}$$

Transforming to the scale of p, we have

$$\begin{aligned}
\text{Power} &= P\left(\hat{p} < p_0 - 2\sqrt{\frac{p_0 q_0}{N}}\,\middle|\, p_1 = \tfrac{6}{10}\right) + P\left(\hat{p} > p_0 + 2\sqrt{\frac{p_0 q_0}{N}}\,\middle|\, p_1 = \tfrac{6}{10}\right)\\
&= P\left(\hat{p} < \tfrac{1}{2} - 2\sqrt{\frac{(\frac{1}{2})(\frac{1}{2})}{N}}\,\middle|\, p_1 = \tfrac{6}{10}\right) + P\left(\hat{p} > \tfrac{1}{2} + 2\sqrt{\frac{(\frac{1}{2})(\frac{1}{2})}{N}}\,\middle|\, p_1 = \tfrac{6}{10}\right)\\
&= P\left(\hat{p} < \tfrac{1}{2} - \frac{1}{\sqrt{N}}\,\middle|\, p_1 = \tfrac{6}{10}\right) + P\left(\hat{p} > \tfrac{1}{2} + \frac{1}{\sqrt{N}}\,\middle|\, p_1 = \tfrac{6}{10}\right)\\
&= P\left[Z < \frac{(\frac{1}{2} - 1/\sqrt{N}) - \frac{6}{10}}{\sqrt{(\frac{6}{10})(\frac{4}{10})/N}}\right] + P\left[Z > \frac{(\frac{1}{2} + 1/\sqrt{N}) - \frac{6}{10}}{\sqrt{(\frac{6}{10})(\frac{4}{10})/N}}\right]\\
&= P\left[Z < \frac{-\sqrt{N} - 10}{\sqrt{24}}\right] + P\left[Z > \frac{-\sqrt{N} + 10}{\sqrt{24}}\right]\\
&= P\left[Z < \frac{-\sqrt{N} - 10}{4.899}\right] + P\left[Z > \frac{-\sqrt{N} + 10}{4.899}\right]
\end{aligned}$$

The easiest way to solve this equation for N is by trial and error methods. This procedure is summarized in Table 12-4.

Table 12-4. Determination of power for testing $p = \frac{1}{2}$ versus $p = \frac{6}{10}$, where $\alpha = \beta = .05$.

Trial	*Value of N*	$\frac{-\sqrt{N} - 10}{4.899}$	$\frac{-\sqrt{N} + 10}{4\ 899}$	*Power*
1	100	−4.08	0	0001 + .5000 = .5001
2	200	−4.93	−.85	0000 + 8023 = .8023
3	300	−5.58	−1.49	0000 + .9319 = .9319
4	324	−5.72	−1.63	0000 + .9484 = .9484

Certainly, a sample size of about 325 to 350 would be recommended for detecting $p > \frac{6}{10}$ with probability of .95. Note that a sample of 200 would be appropriate for a power of .80.

12-5 ONE-SAMPLE TEST FOR THE PARAMETER OF THE HYPERGEOMETRIC DISTRIBUTION

Sometimes the procedure of Section 12-4 may be used to test $H_0: p = p_0$ versus $H_1: p \neq p_0$ if the universe is finite in extent This approximation will be quite adequate if $n/N < 10$ percent. If the sampling fraction exceeds 10 percent, one should use

$$\sigma_{\hat{p}}^2 = \frac{pq}{n}\left(\frac{N-n}{N-1}\right)$$

in place of $\sigma_p^2 = pq/n$ so that the test statistic is given by

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}\left(\dfrac{N-n}{N-1}\right)}}$$

As a use of this test statistic, consider the study of student attitudes toward a school reorganization plan that was reported in Table 3-5. At the time of the survey it was found that about 75 percent of the Caucasian students at School A liked school more. Suppose that a follow-up study were to be conducted the following year, but instead of surveying the entire school only 50 of the 250 students would be sampled. The hypothesis to be tested is that student attitudes had not changed and that p is still equal to .75. Thus, one wishes to test $H_0: p_0 = .75$ versus $H_1: p_1 \neq .75$. Since $n/N = \frac{50}{250} = .20$, the hypergeometric model must be used. Thus

$$\begin{aligned} Z &= \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}\left(\dfrac{N-n}{N-1}\right)}} \\ &= \frac{\hat{p} - \frac{3}{4}}{\sqrt{\dfrac{(\frac{3}{4})(\frac{1}{4})}{50}\left(\dfrac{250-50}{250-1}\right)}} \\ &= \frac{\hat{p} - .75}{.055} \end{aligned}$$

Suppose that in the sample the total number who like school more is given by $T = 40$; then $\hat{p} = \frac{40}{50} = .80$ so that

$$Z = \frac{.80 - .75}{.055} = \frac{.05}{.055} = .91$$

Since $Z < 1.96$, there is no reason to believe that attitudes have changed. Thus, it is still true that $p = .75$.

12-6 ONE-SAMPLE TEST FOR THE VARIANCE OF A POPULATION

In the curriculum study that employed the Penfield Electrical Science Test, it is recalled that the published norms indicate that the variance of the test scores is given by $\sigma^2 = 12^2 = 144$. If one wanted to know whether the variability of the new method was greater or less than that stated in the norms, one could use the following test statistic:

$$\chi^2 = \frac{(N-1)\,S^2}{\sigma_0^2}$$

to test the hypotheses H_0 $\sigma^2 = \sigma_0^2$ versus the alternative H_1: H_0 is false. It should be noted that this same statistical variable was employed for establishing a confidence interval for σ^2.

If the hypothesis is true and if the observations are independent and from a normally distributed population, then the suggested statistic has a chi-square distribution with $\nu = N - 1$ degrees of freedom. In this case, $\nu = 29 - 1 = 28$, so that the 90 percent decision rule is given by

$$\text{D.R.: reject } H_0 \text{ if } \chi^2 < \chi^2_{28}(.05) = 16.928$$
$$\text{or if } \chi^2 > \chi^2_{28}(.95) = 41.337$$

Furthermore, $S = 10.3$, so that the value of the test statistic is

$$\chi^2 = \frac{(29 - 1)(10.3)^2}{(12)^2} = 20.629$$

Thus, there is no reason to doubt the hypothesis that the variance of scores on PEST for the new method is any different from that reported in the manual on norms.

As is recalled, one of the assumptions needed to ensure the chi-square distribution of the statistic is that the parent population be normal in form. In this particular case, this is an extremely important assumption; deviations from it can produce large discrepancies between the actual and nominal type I and type II errors. This is unlike the case for tests of hypothesis and confidence intervals for population centers. For those tests, the normality assumption is less important since the t test procedures are robust with respect to deviations from normality. The chi-square test for variance does not have this property.

12-7 DECISION RULES AND SAMPLE SPACES

In Section 2-13, a sample space was defined as the complete set of possible samples that could be generated by an experiment. Later, in Section 9-2, an experiment was defined as a random selection of a sample point from its corresponding sample space. In Section 8-3 it was seen that a statistic was a number that could be computed from the sample point that was selected and that a statistic possessed a sampling distribution over the entire set of samples that comprised the sample space. Finally, in Section 12-1, a decision rule was defined as those values of the test statistic that lead to rejection of H_0. This means that a decision rule actually defines a subset of the sample space that consists of all possible samples that are not consistent with H_0, the hypothesis under test, but are actually consistent with H_1, the alternative hypothesis.

To clarify these notions, consider the hypothesis-testing model of Section 5-5, in which a test was being run on a woman who claimed to have ESP. As is recalled, the test of the hypothesis that she did not have the ability depended upon the rule to reject the hypothesis of no ability if the number of correct choices in 10 trials was greater than or equal to 8. If 1 corresponds to a correct choice and 0 to an incorrect

choice, then the complete sample space of the study consisted of the 1,024 sample points {(0,0,0,0,0,0,0,0,0,0), (1,0,0,0,0,0,0,0,0,0), (0,1,0,0,0,0,0,0,0,0), . . ., (1,1,1,1,1, 1,1,1,1,1)}. The decision rule of the test said to reject H_0 if X, the number of correct choices, is equal to 8, 9, or 10 The set of outcomes defined by this decision rule consists of the 56 outcomes: {(1,1,1,1,1,1,1,1,0,0), (1,1,1,1,1,1,1,0,1,0), . . . , (1,1,1,1, 1,1,1,1,1,1)}. Thus, it is seen that the decision rule defines the set of outcomes in the sample space, which leads to rejection of H_0. In this sense, the entire sample space consists of outcomes that are compatible either with H_0 or H_1. Furthermore, it is now seen that a type I error occurs when an element of sample space defined by H_0 but consistent with H_1 is generated by an experiment Finally, it is seen that α is the probability of selecting a point of the sample space defined by H_0 that is consistent with H_1.

In a like manner, the decision rule for testing $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ defines a subspace of the sample space that gives sample means consistent with H_1 A type I error occurs if the experiment generates an element of the sample space defined by H_0 that is actually consistent with H_1. In this continuous case, this occurs if Z or t is statistically different from 0.

12-8 CONFIDENCE INTERVALS AND HYPOTHESIS TESTING

From the previous discussion and examples, the similarities between confidence-interval estimation and hypothesis testing should be quite obvious. The assumptions

Table 12-5. One-sample tests and confidence intervals for the parameters of the normal and binomial distributions.

Case	*Hypothesis*	*Test statistic*	*Confidence interval*	*Assumptions*
1	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{N}}$	$\mu = \bar{X} \pm Z\left(\dfrac{\alpha}{2}\right)\dfrac{\sigma}{\sqrt{N}}$	1. Independence 2. Normality 3. Variance known
2	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$t = \dfrac{\bar{X} - \mu_0}{S/\sqrt{N}}$ where $\nu = (N-1)$	$\mu = \bar{X} \pm t_\nu\left(\dfrac{\alpha}{2}\right)\dfrac{S}{\sqrt{N}}$	1. Independence 2. Normality 3. Variance unknown
3	$H_0: p = p_0$ $H_1: p \neq p_0$	$Z = \dfrac{\hat{p} - p_0}{\sqrt{p_0 q_0/N}}$	$p = \hat{p} \pm Z\left(\dfrac{\alpha}{2}\right)\sqrt{\dfrac{\hat{p}\hat{q}}{N}}$	1. Independence 2. Binomial variable 3. $Np > 5$ $Nq > 5$
4	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	$\chi^2 = \dfrac{(N-1)S^2}{\sigma^2}$ where $\nu = (N-1)$	$\dfrac{(N-1)S^2}{\chi_\nu^2(1-\alpha/2)} < \sigma^2 < \dfrac{(N-1)S^2}{\chi_\nu^2(\alpha/2)}$	1. Independence 2. Normality

required for the construction of a confidence interval and the performance of the corresponding statistical test are identical. Furthermore, the decision to accept or reject the hypothesis on the basis of a confidence interval is identical to the decision that one makes from a test of hypothesis. This then poses the question "Why learn two different methods that essentially accomplish the same thing?" The answer to this must be postponed at this time. Suffice it to say that in the one-sample case, it makes no difference. In the two or more sample case, the hypothesis-testing models make it easier to determine *which* confidence intervals are appropriate. In the one-sample case, there is only one confidence interval that may be determined. This is not true for the multisample case.

For those who have not noted the similarities in the two methods, the summaries of Table 12-5 are presented.

12-9 TEST OF HYPOTHESIS FOR A GENERAL PARAMETER

A little reflection on the hypothesis-testing procedures for measures of central tendency indicates that they have certain common properties. What is even more remarkable is that these properties extend to a much larger set of statistical tests Since this was essentially true for the confidence-interval procedures discussed in Chapter 11, it is not surprising that it should be true for associated hypothesis-testing models. For that reason consider the general hypothesis-testing model for testing $H_0\cdot\ \theta=\theta_0$ against the alternative $H_1: \theta \neq \theta_0$. From the universe in which θ is a parameter, let a random sample of size N be selected. From this sample let $\hat{\theta}$ be an estimator of θ and let the sample size be sufficiently large so that the sampling distribution of $\hat{\theta}$ is approximately $N(\theta,\sigma_{\hat{\theta}}^2)$. Under this model the random variable $Z=(\hat{\theta}-\theta_0)/\sigma_{\hat{\theta}}$ can be used as a test statistic for testing the hypothesis. If H_0 is true, this statistic has a distribution that is approximately normal in form with expectation of 0 and variance of 1. Thus, a simple decision rule for rejecting H_0 with the probability of a type I error set equal to α is given by

$$\text{D.R.: reject } H_0 \text{ if } Z < -Z\left(\frac{\alpha}{2}\right) \text{ or if } Z > Z\left(\frac{\alpha}{2}\right)$$

The assumptions for the test are that $\hat{\theta}$ is based upon a random sample from some population of interest and that the sampling distribution of $\hat{\theta}$ is approximately normal with expected value θ and variance $\sigma_{\hat{\theta}}^2$. If the variance of the sampling distribution of $\hat{\theta}$ is unknown, then the standard error of the estimator can be substituted for the variance of $\hat{\theta}$ and the random variable $t=(\hat{\theta}-\theta_0)/\mathrm{SE}_{\hat{\theta}}$ can be used as a test statistic. If H_0 is true, this statistic has an approximate t distribution with ν degrees of freedom associated with the standard error of the estimator. The decision rule for this model is

$$\text{D.R.: reject } H_0 \text{ if } t < t_\nu\left(\frac{\alpha}{2}\right) \text{ or if } t > t_\nu\left(\frac{\alpha}{2}\right)$$

If the alternative hypothesis in either model is that $\theta < \theta_0$ or that $\theta > \theta_0$, then the corresponding decision rule is one-sided. Such a test will be more powerful than one that specifies a two-sided alternative. In any case, the test statistics remain the same, that is, Z if $\sigma_{\hat{\theta}}^2$ is known, t if $\mathrm{SE}_{\hat{\theta}}^2$ is substituted for $\sigma_{\hat{\theta}}^2$.

12-10 THE POWER OF THE GENERAL TEST OF HYPOTHESIS IF $\sigma_{\hat{\theta}}^2$ IS KNOWN

For a one-tailed test of hypothesis in which $\theta_0 < \theta_1$, the hypothesis H_0 is rejected only if $Z > Z(\alpha)$. Thus, the power of the normal distribution test is given by

$$\begin{aligned}\text{Power} &= P[\text{rejecting } H_0 | H_1 \text{ is true}]\\ &= P[Z > Z(\alpha) | H_1 \text{ is true}]\end{aligned}$$

Since the computed Z is given by $Z = (\hat{\theta} - \theta_0)/\sigma_{\hat{\theta}}$, it follows that

$$\begin{aligned}\text{Power} &= P\left[\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} > Z(\alpha) | H_1 \text{ is true}\right]\\ &= P[\hat{\theta} - \theta_0 > Z(\alpha)\,\sigma_{\hat{\theta}} | H_1 \text{ is true}]\\ &= P[\hat{\theta} > \theta_0 + Z(\alpha)\,\sigma_{\hat{\theta}} | H_1 \text{ is true}]\end{aligned}$$

When H_1 is true it follows that $\theta = \theta_1$, so that the actual Z is given by $Z = (\hat{\theta} - \theta_1)/\sigma_{\hat{\theta}}$, and

$$\begin{aligned}\text{Power} &= P[\hat{\theta} - \theta_1 > \theta_0 - \theta_1 + Z(\alpha)\,\sigma_{\hat{\theta}}]\\ &= P\left[\frac{\hat{\theta} - \theta_1}{\sigma_{\hat{\theta}}} > \frac{\theta_0 - \theta_1}{\sigma_{\hat{\theta}}} + Z(\alpha)\right]\\ &= P\left[Z > \frac{\theta_0 - \theta_1}{\sigma_{\hat{\theta}}} + Z(\alpha)\right]\end{aligned}$$

From this last result it can be seen that for a two-tailed test of hypothesis, power can be determined from the following equations:

$$\begin{aligned}\text{Power} &= P\left[\left(Z < -Z\left(\frac{\alpha}{2}\right)\right) \cup \left(Z > Z\left(\frac{\alpha}{2}\right)\right) | H_1 \text{ is true}\right]\\ &= P\left[Z < -Z\left(\frac{\alpha}{2}\right) | H_1 \text{ is true}\right] + P\left[Z > Z\left(\frac{\alpha}{2}\right) | H_1 \text{ is true}\right]\\ &= P\left[Z < \frac{\theta_0 - \theta_1}{\sigma_{\hat{\theta}}} - Z\left(\frac{\alpha}{2}\right)\right] + P\left[Z > \frac{\theta_0 - \theta_1}{\sigma_{\hat{\theta}}} + Z\left(\frac{\alpha}{2}\right)\right]\end{aligned}$$

For most two-tailed tests, one of these two terms will be essentially equal to 0. As a result, one term can usually be ignored and attention can be concentrated strictly on the remaining term. This condition will be apparent in the example of Section 13-1.

If σ_0 is unknown, then one must utilize the t distribution and its related distribution that is generated when $\theta = \theta_1$. If N is large, one can consider the normal distribution as appropriate. The proper small-sample procedures are quite complex and not discussed in this book.

12-11 H_0 VERSUS H_1 OR H_1 VERSUS H_0?

Beginning students frequently have a problem defining H_0 and H_1. This difficulty can be reduced by an understanding of the following simple example. Consider a technician who is working for the local water utility and whose job it is to test the water by counting the number of harmful organisms in samples of water Each time the water is tested a decision must be made as to whether the water is drinkable. Two possible hypotheses for testing are "Water is safe to drink," or "Water is not safe to drink" One of these statements must be H_0 and the other must be H_1

One possible way to test the water is to select N independent samples of water and then test each for the appearance of harmful organisms. If $X =$ number of samples containing harmful organisms, then X is $B(N,p)$, so that the expected number of positive samples is given by $\mu = Np$ Even though $p = 0$ corresponds to the hypothesis that the water is safe to drink, one cannot make direct use of the correspondence since $p = 0$ also corresponds to a degenerate form of the binomial distribution. To circumvent this difficulty, let the hypothesis that water is safe to drink be operationally defined by p being some small number that is less than one in a thousand. With these concessions, one has the following two possible models for hypothesis testing:

Model 1:

$H_0: p \geqslant .001$	Water is not safe
$H_1: p < .001$	Water is safe

Model 2:

$H_0: p < .001$	Water is safe
$H_1: p \geqslant .001$	Water is not safe

One way to determine which one of these models to use is to evaluate the consequences of making type I and type II errors. In model 2, a type I error occurs if it is said that H_1 is true when really H_0 is true. When this occurs, it is concluded that the water is not safe to drink when in reality the water is safe. The consequence of making a type I error in model 2 is a wasting of the taxpayers' money, since the workers at the purification plant would start adding more purifying chemicals to the water than are actually needed. In this same model, a type II error occurs if it is said that H_0 is true when really H_1 is true. This error is associated with the conclusion that the water is safe when in reality the water is not safe. The consequences of making that error are much more serious. In addition to an increase in intestinal

discomfort there is a high probability that some citizens of the community may need hospitalization or else might even die. To reduce the probability of this error, one can make β very small by making N very large, at a considerable cost to the community. Thus, one would not want to test model 2.

In model 1, α measures the probability of concluding that H_1 is true when really H_0 is true; this corresponds to the decision that water is safe when it is really unsafe. This is a very serious error! Fortunately, one can make α, the probability of this serious error, as small as desired without increasing the cost of testing. For model 1, β is the probability of concluding that H_0 is true when really H_1 is true; this corresponds to the decision that water is not safe when it really is safe. The consequence of this error is only wasted money. Thus, the rational point of view would be to use model 1 for testing H_0 versus H_1. So as to give the community maximum protection, one would reject H_0 only if $X = 0$ For all other values of X one would not reject H_0, and so extra purifying chemicals would be added to the water. With this decision rule, $\alpha = (1 - p_0)^N = (.999)^N$. For $\alpha = .001$, N is about 7,000. The power of the test is $P = (1 - p_1)^{7000}$. This is close to 1 for most values of $p < .001$. In any case, when the decision concerning what hypothesis to test is made, one should consider the consequence of making an error, and then control the error that is most costly.

12-12 SUMMARY

In this chapter, an introduction to the principles of statistical hypothesis testing was presented for the one-sample problem As will be noted, these same principles apply to the two- and K-sample problems that are discussed in the following chapters. In all of these models one has a hypothesis that is to be tested for truth value. If evidence is brought forth that is contrary to the hypothesis, it is then rejected. If the evidence is in support of the hypothesis, it is not rejected. In either case, an error may occur. If the hypothesis being tested is true and is incorrectly rejected, it is said that a type I error has occurred. On the other hand, if the hypothesis is not rejected when it is false, it is said that a type II error has occurred. Naturally, the final goal that one has in mind when constructing a statistical test is that of producing a rule of rejection that minimizes the probabilities of making these errors. Unfortunately, such rules are almost nonexistent, except in cases when an entire universe of elements can be measured or observed. Since research is frequently dependent upon samples taken from a universe, the possibilities of making these errors are indeed real.

If H_0 represents the hypothesis under test, H_1 the alternative hypothesis that will be declared true if H_0 is denied, α = probability of a type I error, and β = probability of a type II error, then one can summarize the model as shown in Tables 12-2 and 12-3, and summarized in Table 12-6.

Since α and β cannot be reduced to 0, it is customary to set α equal to some preselected value, usually .05 or .01. If the least favorable outcome for the alternative hypothesis can be predicted, then one can guarantee any preselected value for β, usually .20, .10, .05, or .01, by using the trial-and-error methods described in

Table 12-6. Hypothesis-testing model.

Decision regarding H_0	*Truth value of hypothesis under test*	
	H_0 IS TRUE	H_0 IS FALSE
Reject H_0	$\alpha = P$[type I error]	$1-\beta = P$[correct decision]
Do not reject H_0	$1-\alpha = P$[correct decision]	$\beta = P$[type II error]

Section 12-4 for determining the appropriate sample size. If the value of the parameter cannot be guessed at when H_1 is true, then one should use as large a sample as is economically feasible.

Tests on measures of central tendency are based on the normal or t distribution. Thus, to test $H_0 \cdot \theta = \theta_0$ versus the alternative $H_1: \theta \neq \theta_0$, where θ is a measure of center, one uses $Z = (\hat{\theta} - \theta_0)/\sigma_{\hat{\theta}}$ or $t = (\hat{\theta} - \theta_0)/\mathrm{SE}_{\hat{\theta}}$, depending upon whether $\sigma_{\hat{\theta}}$ is known or unknown. If $\sigma_{\hat{\theta}}$ is known, Z is used and H_0 is rejected if $Z < Z(\alpha/2)$ or if $Z > Z(1-\alpha/2)$. If $\sigma_{\hat{\theta}}$ is unknown, t is used and H_0 is rejected if $t < t_\nu(\alpha/2)$ or if $t > t_\nu(1-\alpha/2)$, where ν is the number of degrees of freedom associated with the $\mathrm{SE}_{\hat{\theta}}$.

Thus, to test $H_0 . \mu = \mu_0$ versus the alternative $H_1: \mu \neq \mu_0$ one uses

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{N}} \qquad \text{or} \qquad t = \frac{\bar{X} - \mu_0}{S/\sqrt{N}}$$

depending upon whether σ or S is employed in the test statistic. To test $H_0: p = p_0$ versus the alternative $H_1: p \neq p_0$, one uses

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0/N}}$$

or its equivalent form

$$Z = \frac{T - Np_0}{\sqrt{Np_0 q_0}}$$

All three tests assume random samples from their parent universe. The tests for expected values also assume that the underlying variable is normal in form. However, this assumption can be relaxed when the sample size is large since the central limit theorem property of means can be relied upon. In the binomial case, it must be assumed that the normal approximation for T is adequate. Generally, this will be the case if $Np > 5$ and $Nq > 5$. If these inequalities are not satisfied, then the procedure described in Section 5-5 must be employed. Finally, if the universe is

finite and if the ratio of the sample size to the universe is larger than 10 percent, then one must use the hypergeometric variance for $\hat{p}$, which is given by

$$\sigma_{\hat{p}}^2 = \frac{pq}{n}\frac{N-n}{N-1}$$

where n equals sample size and N represents the universe size.

Tests of hypotheses about variances are based on the chi-square distribution. The test statistic for H_0: $\sigma^2 = \sigma_0^2$ versus H_1: $\sigma^2 \neq \sigma_0^2$ is given by $\chi^2 = (N-1)S^2/\sigma^2$, which when H_0 is true has a chi-square distribution with $\nu = N-1$ degrees of freedom The assumptions for this test are random sampling and normality of the underlying variable. While the t test is robust to deviations from normality, the chi-square test of variances does not share this property and for this reason must be used with more caution than one would use on tests concerning averages.

EXERCISES

12-1. Define the following terms and give an example of each.

(*a*) Experimental hypothesis
(*b*) Statistical hypothesis
(*c*) Hypothesis under test
(*d*) Alternative hypothesis
(*e*) Type I error
(*f*) Type II error
(*g*) Test statistic
(*h*) Decision rule
(*i*) Power
(*j*) Confidence interval

***12-2.** In Exercise 6-7, what is the conclusion if only 10 youths are returned to court for car stealing for the month in which $16 + 24 + 33 + 27 = 100$ were arrested?

***12-3.** Three years following the study summarized in the table of Exercise 6-1, a random sample of 38 students were sampled. At that time, it was learned that the average number of movies attended in one month was given by $\bar{X} = 5.2$. Is this consistent with the earlier study? What have you assumed in making this decision? Are these assumptions reasonable? Explain.

***12-4.** Answer the last question of Exercise 6-5 on the basis of the hypothesis-testing model presented in this chapter.

***12-5.** Answer Exercise 5-7*d* on the basis of the hypothesis-testing model presented in this chapter.

***12-6.** Answer Exercise 8-10*b* on the basis of the hypothesis-testing model presented in this chapter.

***12-7.** Answer the last question of Exercise 9-4 on the basis of the hypothesis-testing model presented in this chapter.

***12-8.** (*a*) What is the hypothesis tested in the study on IQ described in Section 1-2 and Exercise 9-5?
(*b*) What is the alternative hypothesis?
(*c*) What is the test statistic?
(*d*) What is the decision rule for $\alpha = .05$?
(*e*) What is the decision if the number with IQ below 100 is equal to 37?

***12-9.** In Exercise 12-8, above, what is the probability of the decision rule to reject the hypothesis if the number of children in the sample of 50 with an IQ below 100 exceeds 35? Since the value of p is unknown, the evaluation of this decision rule depends upon assigning various values to p and then determining the power of the test. Perform these power determinations for $p = .5, .6, .7, .8, .9$, and 1.0.

***12-10.** Re-do the one-sample confidence interval Exercises 11-1 and 11-8 as hypothesis-testing models.

13 HYPOTHESIS TESTING IN THE TWO-SAMPLE MODEL

CIGARETTE TESTS SHOW MORE TAR

Washington. The Federal Trade Commission reported yesterday that new laboratory tests had shown significant increases in the tar and nicotine content of many cigarettes.

Comparing the new test results with an October report on earlier tests, the commission said that 32 varieties had increased in tar and 78 in nicotine.

All the reported changes were called "statistically significant." That is, the increase or decrease in each case was at least twice as great as the standard deviation, or allowable error, in the tests.

Commission officials declined to speculate on the cause of the changes. Some experts said they might be due to changes in the tobacco crop. The cigarette industry challenged the accuracy of the tests.

13-1 TEST OF HYPOTHESIS FOR THE DIFFERENCE OF TWO EXPECTED VALUES (VARIANCES KNOWN)

Consider two random variables X_1 and X_2 with unknown expected values μ_1 and μ_2. However, let it be assumed that σ_1^2 and σ_2^2 are known but not necessarily equal. Let independent random samples of size N_1 and N_2 be selected from each population. Let $\theta = \mu_1 - \mu_2$ be a measure of the difference that exists in the two populations with respect to centers. As was shown in Section 11-7, an unbiased estimate of this difference is given by $\hat{\theta} = \bar{X}_1 - \bar{X}_2$. If the sample sizes are sufficiently large or if the parent populations are approximately normal in form, then $\hat{\theta}$ has a sampling distribution that is approximately normal with $E(\hat{\theta}) = \theta = \mu_1 - \mu_2$ and $\sigma_{\hat{\theta}}^2 = \sigma_1^2/N_1 + \sigma_2^2/N_2$. Thus a test statistic for testing H_0: $\theta = 0$ against $H_1: H_0$ is false is given by

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/N_1 + \sigma_2^2/N_2}}$$

If $Z < Z(\alpha/2)$ or if $Z > Z(1 - \alpha/2)$, then the hypothesis under test should be rejected.

As an example of the use of this test, reconsider the example of Section 11-7 in which 85 boys and 72 girls were given a standardized test for which it is known that $\sigma_B^2 = 18^2 = 324$ and $\sigma_G^2 = 20^2 = 400$. The hypothesis of the study under test is that in the population of boys and girls who might take this test under similar conditions, the average achievement scores of the boys and girls are equal. In statistical form, the hypothesis to be tested is H_0: $\mu_B - \mu_G = 0$ and the alternative hypothesis is $H_1: H_0$ is false. The test statistic for the test is given by

$$Z = \frac{\bar{X}_B - \bar{X}_G}{\sqrt{\sigma_B^2/N_B + \sigma_G^2/N_G}}$$

If H_0 is true, this test statistic has a sampling distribution that is approximately normal in form with expected value of 0 and variance of 1. Thus, an $\alpha = .05$ decision rule for rejecting H_0 is given by

D.R.: reject H_0 if $Z < -1.96$ or if $Z > 1.96$

For these data

$$Z = \frac{36 - 51}{\sqrt{\frac{324}{85} + \frac{400}{72}}} = \frac{-15}{3.06} = -4.90$$

Therefore, the appropriate decision is to reject H_0. Thus, the difference between the population parameters is now known to be different from 0. Since the exact value of this difference is still unknown, one would now compute the appropriate confidence interval for the expected average difference so as to have some basis for further research or action. This confidence interval has been determined in Section 11-7. When interpreting the final result on the basis of the computed interval, it was noted that 0 was not included in the interval, and it was suggested

that this represented a difference in parameter values. Thus, the decision reached using the earlier discussed confidence-interval procedure is exactly the same as the decision reached using the corresponding hypothesis-testing procedure.

While hypothesis testing has a long history associated with behavioral research, it is in some respects less important than interval estimation. To know that a tested hypothesis has been rejected is not too informative, since the exact numerical value or possible set of numerical values for the parameter is unknown If a hypothesis is rejected, examination of the associated confidence interval will always show that 0 is not included in the interval, but more important, the limits of the interval will also indicate a possible set of numerical values for the parameters of interest. In this sense a confidence interval conveys more information about the variables of a study. One of the practical advantages that hypothesis testing has over interval estimation is that hypothesis-testing theory simplifies the power computations of a test, provided that an intelligent idea or guess about the alternative hypothesis is possessed by the researcher. In general, this guess is not available and so the determination of the power is frequently of academic interest. While confidence intervals can be used to aid in selecting the sample size for a study (see Section 11-13), hypothesis-testing procedures, on the other hand, treat the problem in a more practical and sophisticated manner. This can be illustrated by an example.

Suppose in planning our study for this section it is desired to detect with a high probability, given by $(1-\beta)=.90$, any difference in parameter value that exceeds 10 points in either a positive or negative direction. Further, let the probability of a type I error be controlled at $\alpha = .05$ and let equal numbers of boys and girls be used for the study. Since the power desired is to exceed .90, the probability of a type II error is to be controlled at $\beta \leqslant .10$. Since the direction of the difference has not been specified, both positive and negative differences $\theta_1 > 10$ or $\theta_1 < -10$, where $\theta_0 = \mu_B - \mu_G = 0$, are of equal interest. The smallest parameter difference, or least significant difference of interest, is given by $\theta = 10$. If the sample sizes are determined for this least favorable outcome or minimal difference of practical interest and if $\theta_1 > 10$, then the actual power will be greater than .90, since for $\theta_1 > 10$, smaller samples will suffice.

For this example, $\theta_0 = 0$, $N_B = N_G = N$, $\sigma_{\hat\theta}^2 = 324/N + 400/N = 724/N$, $Z(\alpha/2) = -1.96$, and $Z(1-\alpha/2) = 1.96$. Substituting these values into the power equation of Section 12-10, we find the power of the test given by

$$\text{Power} = P\left[Z < \left(\frac{0-10}{\sqrt{724/N}}\right) - 1.96\right] + P\left[Z > \left(\frac{0-10}{\sqrt{724/N}}\right) + 1.96\right]$$

$$= P\left[Z < -10\sqrt{\frac{N}{724}} - 1.96\right] + P\left[Z > -10\sqrt{\frac{N}{724}} + 1.96\right]$$

As before, the appropriate sample size N can be determined by trial-and-error methods. The computations for the various trials are summarized in Table 13-1.

Table 13-1. Determination of the sample size for testing $H_0 = 0$ against H_1 $\theta = 10$, with $\alpha = .05$ and $\beta = .10$.

Trial	*Value of N*	$-10\sqrt{\frac{N}{724}} - 1.96$	$-10\sqrt{\frac{N}{724}} + 1.96$	*Power of the test*
1	50	−4.59	−.67	.0000 + .7486 = .7486
2	75	−5.18	−1.26	.0000 + 8962 = .8962
3	76	−5.20	−1.28	.0000 + 8997 = .8997
4	77	−5.22	−1.30	.0000 + 9032 = .9032

As can be seen, about 76 boys and girls should be used in each group. Since these were approximately the sample sizes used for the study, one would feel that the probability of detecting the difference of interest is quite good. In this case, the observed difference of 15 points was considerably greater than the least interesting significant value of 10 points.

Note that in determining the sample size, the probabilities for one of the rejection sets remains at 0. Since this happens in most power computations, one can simplify the computations by ignoring one side of the power equation.

13-2 TESTS OF HYPOTHESIS FOR THE DIFFERENCE OF TWO EXPECTED VALUES (VARIANCES UNKNOWN BUT ASSUMED TO BE EQUAL)

The statistical test presented in this section is the most frequently performed test of behavioral research. For this test, one assumes that the variances of two universes are unknown but equal in value. The hypothesis tested is that H_0: $\theta = \mu_1 - \mu_2 = 0$ and the alternative is $H_1: H_0$ is false. As was shown in Section 11-7, an unbiased estimate of the difference in the population parameters is given by $\hat{\theta} = \bar{X}_1 - \bar{X}_2$, with the squared standard error of the difference given by:

$$\mathrm{SE}^2_{\hat{\theta}} = \frac{S_p^2}{N_1} + \frac{S_p^2}{N_2} = S_p^2\left(\frac{1}{N_1} + \frac{1}{N_2}\right)$$

where

$$S_p^2 = \frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{(N_1 - 1) + (N_2 - 1)}$$

As a result, a test statistic that can be used to test the hypothesis is given by

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\mathrm{SE}_{(\bar{X}_1 - \bar{X}_2)}} = \frac{(\bar{X}_1 - \bar{X}_2)}{S_p\sqrt{1/N_1 + 1/N_2}}$$

In this form, the test is frequently referred to as the two-sample t test, or Student's two-sample t test.

If the parent populations are normal, or if the sample sizes are sufficiently large so as to induce normality in the distribution of the differences in averages, and if independent random samples are selected from each of the populations, then an α percent decision rule for rejecting H_0 is given by

$$\text{D.R.: reject } H_0 \text{ if } t < t_\nu\left(\frac{\alpha}{2}\right) \text{ or } t > t_\nu\left(1 - \frac{\alpha}{2}\right), \text{ where } \nu = (N_1 - 1) + (N_2 - 1)$$

As an example of the use of this statistical test, consider the following hypothetical study in which 9 boys and 13 girls were observed by 5 judges who independently scored each of the children with respect to the degree of hostility they expressed toward one another during an hour play period in a nursery school class. From related research it is reasonable to expect boys to be more aggressive than girls. Thus, a reasonable experimental hypothesis is that boys will show more hostility in play than will girls. Transforming these statements into statistical form, we see that the hypothesis of interest is given by $H_0 \cdot \theta = \mu_B - \mu_G = 0$ and the alternative is given by $H_1: \theta = \mu_B - \mu_G > 0$.

The final criterion variable measured on each child and used to test this hypothesis was the grand total of the individual scores assigned by the five judges for each child. As is known from Theorem 7-8, sums of random variables tend to a normal distribution, even when the parent population is not exactly normal. Thus, it would appear that the normality assumption is reasonably satisfied. The assumption of independence is, however, a bit questionable even though each child is scored independently by the five judges, since it seems reasonable to expect outward hostility to be returned by more hostility. Thus, if one boy is hostile toward one of his playmates, it is reasonable to expect the put-upon child to return the hostility with more hostility. As a result, the conclusions of this study must be tempered with the knowledge that conditions for a valid statistical test are not completely satisfied. Concerning the assumption of equal variance, one might think that the variance for the boys might be larger than the variance for the girls, since boys show greater variability in this trait than do girls. If there is a difference in variances, this test should not be used. If the variances are equal, then t is a reasonable test statistic. With

$$t = \frac{(\bar{X}_B - \bar{X}_G)}{\sqrt{S_p^2/N_B + S_p^2/N_G}}$$

the $\alpha = .05$ decision rule is: D.R.: reject H_0 if $t > t_{20}(.95) = 1.725$. The total scores and sample statistics for the two sets of children are summarized in Table 13-2.

The pooled estimate of the common variance is given by

$$\begin{aligned} S_p^2 &= \frac{(N_B - 1)\, S_B^2 + (N_G - 1)\, S_G^2}{(N_B - 1) + (N_G - 1)} \\ &= \frac{8(487.61) + 12(334.10)}{8 + 12} \\ &= 395.50 \end{aligned}$$

Table 13-2. Total scores and sample statistics for the two samples of children of the study.

Scores of the boys	*Scores of the girls*
82	62
51	28
98	45
80	72
63	68
45	35
112	27
86	79
93	42
	29
	31
	35
	39
Average 78.89	45 54
Variance 487.61	334 10

Thus

$$t = \frac{(\bar{X}_B - \bar{X}_G)}{\mathrm{SE}_{(\bar{X}_B - \bar{X}_G)}} = \frac{(78.89 - 45.54)}{\sqrt{395.50(\frac{1}{9} + \frac{1}{13})}} = \frac{33.35}{8.62} = 3.87$$

Since $t = 3.87 > 1.725$, the hypothesis of no difference in expected hostility scores as measured by the five judges is rejected. The corresponding 95 percent confidence interval for the one-tailed difference is given by

$$\theta = \mu_B - \mu_G > (\bar{X}_B - \bar{X}_G) - t_\nu(.05)\sqrt{S_p^2\left(\frac{1}{N_B} + \frac{1}{N_G}\right)}$$

For the observed data,

$$\mu_B - \mu_G > 33.35 - 1.725(8.62) = 18.48$$

Thus, the mean hostility score for the boys is at minimum 18 points higher than that for the girls.

13-3 TEST OF HYPOTHESIS FOR THE DIFFERENCE OF TWO EXPECTED VALUES (VARIANCES UNKNOWN AND NOT NECESSARILY EQUAL)

In Section 11-9 it was seen that the confidence interval for the difference between two expected values when the variances were unknown and unequal was based

upon the Welch–Aspin approximation to the t distribution. As would be expected, the test statistic for the same model treated in a hypothesis-testing fashion is also based upon this distribution. Thus, the test statistic for testing $H_0\ \theta = \mu_1 - \mu_2 = 0$ versus the alternative H_1: H_0 is false is given by

$$t^* = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{S_1^2/N_1 + S_2^2/N_2}}$$

which, when H_0 is true, is approximately t with ν^* degrees of freedom, where ν^* is as defined in Section 11-9.

As an example of the use of this test, consider a hypothetical archaeological investigation in which 8 hand-cutting tools were found near a stream and 15 other similar tools were found in a region 2 miles away. At the time of the finding it was of interest to know whether the tools were designed to satisfy different needs and functions related to hunting or food processing. In answering this question, a number of measures were taken on each specimen. One of the variables measured was the length of the longest axis. If the hand tools were made for identical functions, then it might be expected that the average lengths would be about the same. Thus, a hypothesis of possible interest is H_0: $\mu_1 - \mu_2 = 0$ versus the alternative H_1: H_0 is false. The statistics for the cutting tools are summarized in Table 13-3. Assuming

Table 13-3. Statistics for the two sets of hand tools.

Sample statistics	*Population 1. Sample found near stream*	*Population 2. Sample found away from stream*
Sample size	8	15
Average	7.8 cm	8.2 cm
Variance	3.6 cm²	.8 cm²

that these fortuitously found specimens represent random samples of hand tools for the two populations of different tool functions, it appears that the variances in the hypothetical universes of hand-tool length are quite different. Thus, the Welch–Aspin test is appropriate. The computations are summarized as they were for the confidence-interval procedure.

1. $\hat{\theta} = \bar{X}_1 - \bar{X}_2 = 7.8 - 8.2 = -.4$
2. $\nu_1 = N_1 - 1 = 8 - 1 = 7$
3. $\nu_2 = N_2 - 1 = 15 - 1 = 14$
4. $\text{SE}^2_{\bar{X}_1} = \dfrac{S_1^2}{N_1} = \dfrac{3.6}{8} = .4500$

5. $SE^2_{\bar{X}_2} = \dfrac{S_2^2}{N_2} = \dfrac{.8}{15} = .0533$

6. $SE^2_{\bar{X}_1} + SE^2_{\bar{X}_2} = .4500 + .0533 = .5033$

7. $C = \dfrac{.4500}{.5033} = .8941$

8. $\nu^* = \dfrac{(7)(14)}{14(.8941)^2 + 7(.1059)^2} = 8.70$

   Since ν^* rarely equals an integer value, one is advised to reduce ν^* to the next lowest integer value This has the effect of making the test more conservative in that the probability of a type II error is increased. Thus, $\nu^* = 8$.

9. Since the alternative hypothesis is two-sided, this is a two-tailed test and H_0 should be rejected if $t^* < t_8(.025) = -2.31$ or if $t^* > t_8(.975) = 2.31$. Note how the apparent degrees of freedom have been reduced from 21 to 8. This means that the power of the resulting t^* test is less than expected. When the variances are quite different this reduction is reasonable.

10. $t^* = \dfrac{-.4}{\sqrt{.5033}} = \dfrac{-.4}{.7094} = -.56$

11. H_0 is not rejected. On the basis of the tool lengths it would be concluded that the tools could have been made by the Indians for the same function.

Since the sample variances are so different, and since the samples are not random samples, there is further reason to wonder or doubt the decision. These hand tools were produced a good number of years ago and their finding at the two points is dependent upon an archaeologist searching at the spot where "nature" protected the man-made cutting instruments from the elements. As a result it may be difficult to believe that these hand tools represent a true random sample from the population of hand tools. Consequently, the conclusion of nondifferences in averages should be interpreted with caution.

This is a very common problem of observational studies in which the researcher has not deliberately manipulated the external conditions of the investigation. This is not to say that observational studies are less desirable than planned experimental studies or that one cannot make valid inferences from observational studies. Instead, it only means that the researchers who conduct observational studies have a greater responsibility to be attuned to the assumptions that are required for certain statistical tests employed in the analysis. Furthermore, they must also be prepared to limit their conclusions if the assumptions for the statistical tests employed are clearly violated. While observational studies are generally more difficult to execute than experimental studies, they are in some respects the most important kinds of studies

that the behavioral scientist can conduct. The questions that are asked in observational studies are generally of greater practical and theoretical importance than are the questions asked in planned experimental studies. This is especially true of field and survey research studies in anthropology, education, political science, public health, sociology, and social welfare. This statement is less true for experimental studies in biology, medicine, psychology, and zoology, even though they have their unique problems.

13-4 TEST OF HYPOTHESIS FOR THE DIFFERENCE OF TWO EXPECTED VALUES WHEN THE SAMPLES ARE NOT STATISTICALLY INDEPENDENT

When the members of one sample are paired with the members of another sample or if the same sample members are observed in a before-and-after experiment, the outcomes on each member of a pair are not statistically independent. As was shown in Section 11-10, one could handle this problem by converting the two-sample problem to a one-sample problem, taking the differences in observed values for the paired members. Thus, if d represents the difference between the paired observations, then the hypothesis to be tested is that $H_0 \cdot \theta = \mu_d = \mu_1 - \mu_2 = 0$ versus the alternative $H_1 \cdot H_0$ is false.

In Section 11-10 it was seen that an unbiased estimate of the parameter difference is given by $\hat{\theta} = \bar{d}$ and the standard error of the difference is given by $SE_{\bar{d}} = S_d/\sqrt{N}$. Thus, a test statistic for testing the hypothesis of interest is given by

$$t = \frac{\bar{d}}{S_d/\sqrt{N}} = \frac{\bar{d}\sqrt{N}}{S_d}$$

which, when H_0 is true, is approximately t with $\nu = N - 1$ degrees of freedom.

To illustrate this test, consider the following hypothetical study. The school records of the 10 high schools of a large school district were searched until 20 senior boys were found who had sisters who were also seniors in school and were within one year of age of their brother. The members of each pair were asked to record the amount of money spent on amusements and luxuries such as records, cigarettes, dates, cokes, sundaes, dances, movies, magazines, gasoline, etc., for the complete month of March in 1965. The amount of money spent by each student rounded to the nearest dollar is reported in Table 13-4. As can be seen, if the male member of a pair spent large sums of money on entertainment, his sister also spent large sums. If the male member spent small sums, his sister also spent small sums. The amount of spending by pairs is said to be correlated The existence of this correlation is not unusual, since one would expect the amount of money that brothers and sisters in the same household have for spending to be about the same unless one of the pairs has a part-time job after school or on weekends. This would be especially true for the boys. Furthermore, when a boy dates a girl, he also picks up the tabs; his date goes free. Thus, a reasonable experimental hypothesis is that boys spend more money on entertainment than girls. The hypothesis to be tested is that the amount

Table 13-4. Amount of money spent on amusements in March 1965.

Pair	*Amount spent by male member*	*Amount spent by female member*	*Difference*
1	41	27	14
2	19	18	1
3	22	11	11
4	10	28	−18
5	17	11	6
6	9	14	−5
7	19	20	−1
8	11	21	−10
9	32	17	15
10	7	7	0
11	68	48	20
12	20	14	6
13	34	19	15
14	76	65	11
15	34	28	6
16	45	35	10
17	73	70	3
18	69	61	8
19	28	22	6
20	32	29	3

of money spent on entertainment is unrelated to the sex of the spender. Stated in statistical form, $H_0 \cdot \theta = \mu_d = \mu_B - \mu_G = 0$ versus the alternative hypothesis $H_1: \theta = \mu_d = \mu_B - \mu_G > 0$. The test statistic is given by $t = \bar{d}\sqrt{N}/S_d$. With $\alpha = .05$, H_0 should be rejected if $t > t_{19}(.95) = 1.729$.

The difference scores are listed in Table 13-4. The average difference value is equal to

$$\bar{d} = \frac{\sum_{i=1}^{20} d}{N} = \frac{101}{20} = \$5.05$$

and the variance of the difference value is equal to

$$S_d^2 = \frac{N \sum_{i=1}^{20} d_i^2 - \left(\sum_{i=1}^{20} d_i\right)^2}{N(N-1)} = \frac{20(2065) - (101)^2}{20(19)} = 81.84$$

The value of the test statistic is given by

$$t = \frac{5.05}{\sqrt{81.84/20}} = \frac{5.05}{2.02} = 2.50$$

Thus, there is reason to reject H_0. It is concluded that boys do spend more money on entertainment than do their sisters at the same age and grade level. The one-sided 95 percent confidence interval for the difference between the two parameters is given by

$$\begin{aligned} \theta = \mu_B - \mu_G &> \bar{d} - t_\nu(.05)\,\mathrm{SE}_{\bar{d}} \\ &= 5.05 - 1.729(2.02) = \$1.55 \end{aligned}$$

Many people would agree that a difference as small as \$1.55 is not of much practical importance even though it is statistically important. This, by the way, is a frequent finding in many behavioral studies. The differences in parameter values may be statistically significant but when examination is made of the confidence interval, the observed difference may be seen to be of minor importance. This points up another advantage of interval estimation over hypothesis testing. Statistical differences can be evaluated for meaningful differences.

In this example, the assumption is made that the differences are approximately normal in form or that the number of pairs is sufficiently large so that the sampling distribution of $\bar{d}$ is approximately normal in form. For this example, this assumption is not unreasonable since a single pair difference is of the general form $d_i = (+1)\,X_{B_i} + (-1)\,X_{G_i}$, which tends to have an approximately normal distribution. Furthermore, since $\bar{d}$ is an average of these kinds of variables, it probably has a normal or near-normal form also.

For this test it is further assumed that the spending of money between pairs is independent. In a large community this is not an unreasonable assumption. Finally, it must be assumed that the 20 pairs of brother and sister combinations represent a random sample of such brother-sister pairs in the universe of students who attend the senior classes of these schools. This assumption is doubtful since these 20 pairs were found by searching the school records, yet, they might truly represent the entire collection of such pairs. In any case, to use this test one must argue that the 20 pairs are a random sample of brother-sister pairs that could pass through the school system. In this sense, the population of the study is abstract and purely hypothetical. Frequently, this is the only way that one can justify statistical testing procedures. In every research study, the behavioral scientist should decide whether the population to which inference is to be made is real or hypothetical. Generally, it will be hypothetical.

In Section 5-6, the sign test was presented as an alternative test procedure. While the sign test is easier to perform, it is in general not recommended since it is less powerful than the matched pair t test. However, if the distribution of differences does not appear normal in form, consideration should be given to the sign test as an alternative procedure.

13-5 TESTS OF HYPOTHESIS FOR THE DIFFERENCE OF THE PARAMETERS OF TWO BINOMIAL VARIABLES

Next to the two-sample t test, the statistical test of this section is one of the most frequently performed tests of behavioral research. For this test let X_1 be $B(N_1,p_1)$ and let X_2 be $B(N_2,p_2)$ and let these variables be statistically independent. As was seen in Section 11-11, an unbiased estimate of the difference between the two binomial parameters is given by $\hat{\theta}=\hat{p}_1-\hat{p}_2$ and the squared standard error of the difference is given by

$$\mathrm{SE}^2_{\hat{\theta}}=\frac{\hat{p}_1\hat{q}_1}{N_1}+\frac{\hat{p}_2\hat{q}_2}{N_2}$$

This would suggest that one could create a test statistic by taking

$$Z=\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}_1\hat{q}_1/N_1+\hat{p}_2\hat{q}_2/N_2}}$$

and rejecting $H_0\cdot\ \theta=p_1-p_2=0$ versus the alternative $H_1\cdot\ H_0$ is false, if $Z<Z(\alpha/2)$ or if $Z>Z(1-\alpha/2)$. This procedure would not be appropriate because the standard error of the difference, as defined in Section 11-11, is for the difference between two estimates of different p values. In this case, a standard error for the difference of two p values, in which it is hypothesized that the p values are identical and not different, is needed. When $p_1=p_2=p_0$, then the variance of the difference in two estimated p values is given by

$$\sigma^2_{\hat{p}_1-\hat{p}_2}=\frac{p_0q_0}{N_1}+\frac{p_0q_0}{N_2}$$

and the squared standard error of the difference is given by

$$\mathrm{SE}^2_{\hat{p}_1-\hat{p}_2}=\frac{\hat{p}_0\hat{q}_0}{N_1}+\frac{\hat{p}_0\hat{q}_0}{N_2}$$

Since $\hat{p}_0$ is an average of 1's and 0's, an unbiased estimate of it is given by

$$\hat{p}_0=\frac{N_1\hat{p}_1+N_2\hat{p}_2}{N_1+N_2}=\frac{X_1+X_2}{N_1+N_2}$$

where X_1 and X_2 equal the number of 1's in the corresponding samples. With this, one can now create a test statistic for testing H_0. The test statistic is given by

$$Z=\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}_0\hat{q}_0/N_1+\hat{p}_0\hat{q}_0/N_2}}$$

As an example of the use of this test, consider the following hypothetical study. In a medium-sized city the school administration was concerned about the differential drop-out rate that existed in the two senior classes of the two high schools of the community. As a result, it was decided to study the matter in detail. As a part of the

study, it was decided to determine whether there was a "real" difference in the drop-out rate for the two schools. For this study, the 1962–1963 school year records were employed. It was argued that the class of 1962–1963 was similar to the school populations that had been going through the schools during the last few years, and therefore these two classes could be treated as though they were random samples of students that normally pass through the school. As to whether this is a valid argument is a moot point, but if one argues that the forces that operate on students in bringing them to the school at the time of the study are random in nature, then maybe the argument is not unrealistic.

The hypothesis to be tested is that the drop-out rates for the two schools are equal in value versus the alternative that they are unequal. Stated in statistical terms, this hypothesis and its alternative are H_0: $\theta = p_1 - p_2 = 0$ and $H_1 \cdot H_0$ is false. The test statistic is given by

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0 \hat{q}_0 / N_1 + \hat{p}_0 \hat{q}_0 / N_2}}$$

With $\alpha = .05$, the hypothesis should be rejected if $Z < -1.96$ or if $Z > 1.96$. After the records were collected, examined, and tabulated, the statistics for the 1962–1963 school year were computed and listed as shown in Table 13-5.

Table 13-5. Statistics for the two schools in 1962-1963.

State of student at end of year	*School 1*	*School 2*	*Total*
Graduated	193	541	734
Did not graduate	38	41	79
Total	231	582	813

The conditional and unconditional drop-out rates for the schools were as follows:

$$\hat{p}_1 = \frac{X_1}{N_1} = \frac{38}{231} = .1645$$

$$\hat{p}_2 = \frac{X_2}{N_2} = \frac{41}{582} = .0704$$

$$\hat{p}_0 = \frac{X_1 + X_2}{N_1 + N_2} = \frac{79}{813} = .0972$$

The value of the test statistic is given by

$$Z = \frac{1645 - .0704}{\sqrt{(\ 0972)(\ 9028)/231 + (.0972)(.9028)/582}}$$

$$= \frac{.0941}{.0230}$$

$$= 4.09$$

As a result, the hypothesis is rejected and it is concluded that the drop-out rates for the two schools are different. The 95 percent confidence interval for the difference in drop-out rates is given by

$$(p_1 - p_2) = (\hat{p}_1 - \hat{p}_2) \pm Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1} + \frac{\hat{p}_2\hat{q}_2}{N_2}}$$

$$= (.1645 - .0704) \pm 1.96\sqrt{\frac{(.1645)(.8355)}{231} + \frac{(.0704)(.9296)}{582}}$$

or

$$.0420 < p_1 - p_2 < .1462$$

Note that the decision reached here is a bit strained; about the only way one can justify it is to make the argument that the class of students in 1962–1963 is really a random sample of students that pass through the school over time. By doing this one assumes that student populations are static and that the same kind of student is constantly entering the school. As is well known, human populations are perpetually in flux and change. As a result, one must fall back on assuming some hypothetical population of students for the two schools. Unless one can perform these mental gymnastics, one should not be using statistical tests in situations similar to the one of this example. If one decides to make these mental gymnastics, then one should be prepared to temper one's decisions accordingly.

As mentioned earlier, this is a constant problem in most observational studies that are conducted by behavioral scientists. Generally, the independent variables such as sex, race, or yearly income have not been directly manipulated by the researcher. Instead, "nature" is permitted to manipulate the environment. Frequently, this means that undetected sources of bias may enter the study and confound and confuse the results. While the experimentalist is on somewhat stronger grounds when making statistical inferences, the behavioral scientist who works with observational data must always be on his guard and must always be prepared to explain unusual outcomes. Frequently it is difficult to defend the assumption of random sampling with observational studies, and for this reason it might be necessary to qualify the conclusions that one can draw from such studies. Each study has its own unique complications. For that reason, researchers are advised to consult with professional statisticians when problems of sampling and interpretation arise.

Generally, this aid should be sought before the study is begun and not after the data have been collected It is easier to salvage a study when it is planned. Generally, one can do little when it is over.

13-6 CONDITIONAL PROBABILITIES AS MEASURES OF RELATIONSHIP

In Section 3-10 it was stated in an unqualified manner that all research is devoted to estimating conditional probabilities and testing hypotheses in order to determine whether they are equal to unconditional probabilities. For the most part, this statement has been defended throughout the previous discussion. In particular, the test of hypothesis H_0: $p_1 = p_2 = p_0$ has rested on the validity of this broadly based assumption

Thus, in the example of Section 13-5, as soon as the hypothesis of no difference in drop-out rates was rejected, it was immediately known that the confidence interval for $p_1 - p_2$ would not cover 0. In addition, it was known that $\hat{p}_0 = .0972$ could not be used to describe the drop-out rates at the two schools. Instead, it was necessary to report $\hat{p}_1 = .1645$ at School 1, while $\hat{p}_2 = .0704$ at School 2. Thus, rejection of H_0 also implied that the conditional probabilities of drop-out are unequal at the two schools.

In like manner, when it was noted that the confidence interval of $p_W - p_{\bar{W}}$ of Table 11-3 for the difference in percent agreement for whites and nonwhites concerning a question about the changing of school boundaries did not include 0, it was immediately seen that $\hat{p}_0$ was not an appropriate measure when discussing agreement to school boundary changes. Instead, recourse had to be made to the estimate $\hat{p}_W = 514$ when discussing attitudes of the whites, while when discussing attitudes of the nonwhites, $\hat{p}_{\bar{W}} = .836$ became important. In this sense, it could be inferred that about 50 percent of the whites supported the recommendation while about 85 percent of the nonwhites also were supporters. In any case, there were large racial differences in the amount of support given to the changing of school boundaries to achieve school integration.

While it is not so apparent, the same statements about conditional and unconditional probabilities also apply to the test of H_0: $\mu_1 = \mu_2 = \mu_0$. For example, if the t test leads to a rejection of $\mu_1 = \mu_2$, then it immediately follows that corresponding conditional probabilities are not equal. Thus, if $\mu_1 \neq \mu_2$, then it must be true that for any X_0,

$$P(X > X_0 | \mu_1) \neq P(X > X_0 | \mu_2)$$

In like manner, the same statement follows for an analysis-of-variance hypothesis rejected by the F test.

Thus, for the data of Table 13-2, when it was seen that the hypothesis of equal expected values in the hostility scores was rejected, then one was compelled to focus on the individual sample means in discussing the findings. For the boys the mean hostility score is given by $\bar{X}_B = 78\ 9$, while the mean hostility score for the

girls is given by $\bar{X}_G = 45.5$ If one knew the exact value of σ^2, one could compute conditional probabilities of interest. Since $S_p^2 = 395\ 50$, one could behave as though $\sigma^2 = 400$ or $\sigma = 20$. If average hostility is equivalent to $X = 50$, then

$$\hat{P}(X > 50|\text{boy}) = \hat{P}\left(Z > \frac{50 - 78.9}{20}\right) = \hat{P}(Z > -1\ 445) = \ 925$$

while

$$\hat{P}(X > 50|\text{girl}) = \hat{P}\left(Z > \frac{50 - 45.5}{20}\right) = \hat{P}(Z > \ 275) = .39$$

In this case it would be inferred that about 90 percent of the boys express above-average hostility while 40 percent of the girls also express above-average hostility, at least as measured by the judges' scoring of their behavior in free play.

13-7 TWO-SAMPLE TESTS AND CONFIDENCE INTERVALS

In Table 13-6, the association between two-sample tests and two-sample confidence intervals is presented for the analysis of this chapter As was done in Table 12-5, all one-sided tests are presented as two-sided tests, mainly to conform to the actual two-sample cases

13-8 SUMMARY

In this chapter, two-sample tests of hypothesis for measures of central tendency were presented and compared to the corresponding confidence-interval procedures. Except for the two-sample binomial case, the procedures are essentially the same and the decisions reached are exactly identical.

The test statistic that one ultimately employs to test H_0: $\mu_1 = \mu_2$ depends upon the assumptions that a researcher is willing to make about the universe under investigation and about the method of sampling employed. If independent random samples are selected from universes in which the variances are known, the appropriate test statistic is given by

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/N_1 + \sigma_2^2/N_2}}$$

which, when H_0 is true, is distributed as an $N(0,1)$ variable. If the variances are unknown but believed to be equal, the appropriate test statistic is given by

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2/N_1 + S_p^2/N_2}}$$

with

$$S_p^2 = \frac{(N_1 - 1)\, S_1^2 + (N_2 - 1)\, S_2^2}{(N_1 - 1) + (N_2 - 1)}$$

Table 13-6. Hypothesis testing in the two-sample model.

Case	*Hypotheses*	*Test statistic*	*Confidence interval*	*Assumptions*
5	$H_0: \mu_1 = \mu_2$ or $\theta = \mu_1 - \mu_2 = 0$ H_1: H_0 is false	$Z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/N_1 + \sigma_2^2/N_2}}$	$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$	1. Independence between samples 2 Independence within samples 3 Normality 4. Variances are known
6	$H_0: \mu_1 = \mu_2$ or $\theta = \mu_1 - \mu_2 = 0$ H_1: H_0 is false	$t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2/N_1 + S_p^2/N_2}}$ where $S_p^2 = \dfrac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{(N_1 - 1) + (N_2 - 1)}$ $\nu = (N_1 - 1) + (N_2 - 1)$	$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm t_\nu\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_p^2}{N_1} + \frac{S_p^2}{N_2}}$	1. Independence between samples 2. Independence within samples 3 Normality 4 Variance unknown 5. Variances equal
7	$H_0: \mu_1 = \mu_2$ or $\theta = \mu_1 - \mu_2 = 0$ H_1: H_0 is false	$t^* = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/N_1 + S_2^2/N_2}}$ where $\nu = \nu^*$	$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm t_{\nu^*}^*\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}$	1. Independence between samples 2 Independence within samples 3. Normality 4. Variances unknown
8	$H_0 \cdot \mu_1 = \mu_2$ or $\theta = \mu_1 - \mu_2 = 0$ H_1: H_0 is false	$t = \dfrac{\bar{d}\sqrt{N}}{S_d}$ where $\nu = N - 1$	$\mu_1 - \mu_2 = \bar{d} \pm t_\nu\left(\frac{\alpha}{2}\right)\left(\frac{S_d}{\sqrt{N}}\right)$	1 Dependence between samples 2 Independence within samples 3 Normality 4. Variances unknown
9	$H_0: p_1 = p_2$ or $\theta = p_1 - p_2 = 0$ H_1: H_0 is false	$Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0\hat{q}_0/N_1 + \hat{p}_0\hat{q}_0/N_2}}$ where $\hat{p}_0 = \dfrac{N_1\hat{p}_1 + N_2\hat{p}_2}{N_1 + N_2}$	$p_1 - p_2 = (\hat{p}_1 - \hat{p}_2) \pm Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1} + \frac{\hat{p}_2\hat{q}_2}{N_2}}$	1 Independence between samples 2 Independence within samples 3. Binomial 4 $N_1 p_0 > 5$, $N_2 p_0 > 5$ $N_1 q_0 > 5$, $N_2 q_0 > 5$

When H_0 is true, this statistic has a t distribution with $\nu = (N_1 - 1) + (N_2 - 1)$ degrees of freedom. Finally, if it is believed that the unknown variances are unequal, then the appropriate test statistic is given by

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/N_1 + S_2^2/N_2}}$$

which, when H_0 is true, is distributed approximately as a t variable with ν^* degrees of freedom where

$$\nu^* = \frac{(N_1 - 1)(N_2 - 1)}{(N_2 - 1)\,C^2 + (N_1 - 1)(1 - C)^2}$$

with

$$C = \frac{\mathrm{SE}^2_{\bar{X}_1}}{\mathrm{SE}^2_{\bar{X}_1} + \mathrm{SE}^2_{\bar{X}_2}}$$

If the observations in the two samples are not statistically independent but are matched on some external criteria or if repeated measures are taken on the sample elements, then one makes a simple conversion to a one-sample problem by computing difference scores for each element of the sample. After this is completed, the test of $H_0\colon \mu_1 = \mu_2$ is made by computing

$$t = \frac{\bar{d}}{S_d/\sqrt{N}}$$

where $\bar{d}$ equals the average of the differences and S_d equals the standard deviation of the differences. When H_0 is true, this test statistic has a t distribution with $\nu = N - 1$.

One assumption common to all of these tests is that the underlying variable is normally distributed. When the sample sizes are large, the normality assumption can be relaxed, since empirical research has shown that the actual probability of a type I and type II error is not materially different from the stated probability, even when the parent universes are not normal. However, in the small-sample case, the assumption of normality takes on greater weight since the probability of a type II error may be inflated. If this should be the case, the researcher is strongly advised to seek out the help of a statistician or else to employ a nonparametric procedure such as the two-sample Wilcoxin test, which is not discussed in this book.

Most likely, the nicotine and tar comparisons reported by the Federal Trade Commission in the quotation at the beginning of this chapter were based on the use of these two-sample tests with $\alpha = .05$. For large samples, a decision rule for $\alpha = .05$ would essentially state that the hypothesis of no difference should be rejected if the observed difference exceeds 2 standard deviations on the criterion variables Clearly, this was the decision rule employed to compare the outcomes for the study on tobacco characteristics

The test of the hypothesis $H_0: p_1 = p_2$ is based upon the following test statistic:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0\hat{q}_0/N_1 + \hat{p}_0\hat{q}_0/N_2}}$$

with

$$\hat{p}_0 = \frac{N_1\hat{p}_1 + N_2\hat{p}_2}{N_1 + N_2}$$

which, when H_0 is true, has a distribution that is approximately $N(0,1)$. In addition to the binomial assumptions of constant probability of occurrence and independence between and within the samples, one must assume that $N_1 p_0 > 5$, $N_2 p_0 > 5$, $N_1 q_0 > 5$, and $N_2 q_0 > 5$. If these inequalities are not satisfied, one should not use this approximate test but instead should use the exact form of the test, which is called the Irwin–Fisher exact test (discussed in Section 5-9).

EXERCISES

***13-1.** Answer Exercise 11-1*d* as a hypothesis-testing model. State the hypothesis, the alternative, the decision rule, the test statistic, and the conclusion. What assumptions have you made? Are they reasonable? Explain.

***13-2.** Answer Exercise 7-1*c* as a hypothesis-testing model. State the hypothesis, the alternative, the decision rule, the test statistic, and the conclusion. What assumptions have you made? Are they reasonable? Explain.

13-3. In a study of the importance of height on final game score for professional basketball players, the height distribution of a well-known professional team was dichotomized at $X = 6$ feet, 2 inches. For each player the average number of points scored per game, excluding free throws, was determined for the 1967–1968 season. The resulting statistics were as shown in the following table. On the basis of this evidence, is height an important characteristic for basketball players?

Height	*Under 6′ 2″*	*Over 6′ 2″*
Mean points per game	8.6	12.3
Standard deviation	1.6	2.2
Number of players	12	17

***13-4.** Answer Exercise 11-6 as a hypothesis-testing model. State the hypothesis, the alternative, the decision rule, the test statistic, and the conclusion. What assumptions have you made? Are they reasonable? Explain.

***13-5.** Analyze the data of Table 5-8 as a hypothesis-testing model. State the hypothesis, the alternative, the decision rule, the test statistic, and the conclusion. What assumptions have you made? Are they reasonable? Explain.

***13-6.** Analyze the data of Exercise 11-9 as a hypothesis-testing model. State the hypothesis, the alternative, the decision rule, the test statistic, and the conclusion. What assumptions have you made? Are they reasonable? Explain.

***13-7.** Consider the Negro and Caucasian students at School 1 of Tables 3-2 and 3-5. Do they have different attitudes toward school? Also do the analysis for the students at School 2. In each case, state the hypothesis, the alternative, the decision rule, the test statistic, and the conclusion. What assumptions have you made? Are they reasonable? Explain.

***13-8.** The data of Exercise 3-10 represent repeated measures taken on a single group of 100 individuals An interesting question associated with these data is "Are attitudes toward the integration of the elementary schools identical to those toward the integration of the junior high schools?" One way to analyze these data is to superimpose a Likert scale (see Exercise 6-9) on the categories and then do a match pair test, first setting up the computation table on page 327. Once this table is completed, one can test the hypothesis that the average Likert score for the two sets of responses is equal. Perform this test. What is your decision? Since the difference scores are clearly not normal, what can you say about the decision?

***13-9.** Analyze the data of Exercise 13-8 as a sign test (use Table A-2).

***13-10.** The test of $H_0: p_1 = p_2$ versus $H_1 \cdot p_1 \neq p_2$ requires that $N_1 p_0 > 5$, $N_2 p_0 > 5$, $N_1 q_0 > 5$, and $N_2 q_0 > 5$. When one of these conditions is not satisfied, the Irwin–Fisher test should be employed. This test is based on hypergeometric probabilities and is discussed in Section 5-7. As an example of the use of this test, consider a study on the teaching of reading where 14 remedial readers were randomly assigned to a control and experimental condition. At the end of the study, the 14 students were tested and rated for improvement in reading skills The results are as shown in the following table. Test the hypothesis that no differences exist in performance with the experimental and control method.

	Control	*Experimental*	*Total*
Did not improve	1	4	5
Improved	5	4	9
Total	6	8	14

Attitude toward integration		*Likert scale*		*Difference score*	*Frequency*
OF ELEMENTARY SCHOOLS	OF JUNIOR HIGH SCHOOLS	FOR ELEMENTARY SCHOOLS	FOR JUNIOR HIGH SCHOOLS		
Strongly agree	Strongly agree	4	4	0	5
Moderately agree	Strongly agree	3	4	−1	8
Moderately disagree	Strongly agree	2	4	−2	12
Strongly disagree	Strongly agree	1	4	−3	10
Strongly agree	Moderately agree	4	3	1	6
Moderately agree	Moderately agree	3	3	0	10
Moderately disagree	Moderately agree	2	3	−1	8
Strongly disagree	Moderately agree	1	3	−2	14
Strongly agree	Moderately disagree	4	2	2	4
Moderately agree	Moderately disagree	3	2	1	2
Moderately disagree	Moderately disagree	2	2	0	8
Strongly disagree	Moderately disagree	1	2	−1	2
Strongly agree	Strongly disagree	4	1	3	1
Moderately agree	Strongly disagree	3	1	2	2
Moderately disagree	Strongly disagree	2	1	1	1
Strongly disagree	Strongly disagree	1	1	0	7

One year later, the study was repeated, but with modifications and with a decision to perform a one-tailed test of hypothesis in which the alternative hypothesis was H_1. $p_E - p_C > 0$. This time, 20 students were included in the study, and the outcomes were as shown in the following table. What would be the conclusion based on the performance of the Irwin–Fisher test with $\alpha \leqslant .05$?

	Control	*Experimental*	*Total*
Did not improve	1	5	6
Improved	8	6	14
Total	9	11	20

14 THE *F* DISTRIBUTION AND TESTS OF VARIANCE

A study published at Columbia last May in the Teachers College Record shows that the present total cost of public schools to local, State, and Federal governments comes to 53 cents per student per hour, or less than most people pay for baby-sitting.

The average figure conceals a wide dispersion for different school districts. Although New York City spends over $900 per student per year, Alabama's bill is less than $300, and the differences within states are as wide as those among states.

14-1 THE *F* DISTRIBUTION

In Chapter 12, the one-sample tests for expected values and variance were presented. In Chapter 13, the two-sample test for equality of expected values was presented but no mention was made of the corresponding test of equal variances. The reason for this omission is that this test requires further development of statistical theory Fortunately, the appropriate theory is easy to derive and understand, but before the theory for the test can be developed with ease it will be necessary to introduce another probability distribution that is very closely associated with the chi-square distribution. This probability distribution is called the F distribution. It is named in honor of Sir Ronald Fisher, one of the giants of modern statistical methods, who indirectly derived the F distribution and proposed some of the statistical tests of this and the next chapter With this introduction, consider the basic nature of the F distribution as described in the following theorem, which is presented without proof.

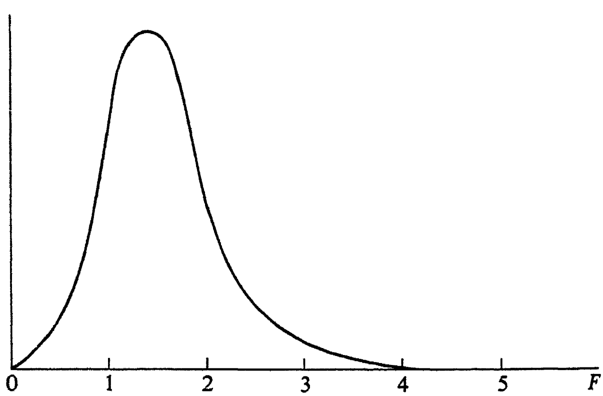
Theorem 14-1

Consider two independent random variables U_1 and U_2. Let U_1 be distributed as a chi-square variable with ν_1 degrees of freedom and let U_2 be distributed as a chi-square variable with ν_2 degrees of freedom. The variable

$$F = \frac{U_1/\nu_1}{U_2/\nu_2}$$

which is simply the ratio of two independent chi-square variables, each divided by its respective degrees of freedom, has an F distribution with ν_1 and ν_2 degrees of freedom.

Figure 14-1. The F distribution with $\nu_1 = 7$ and $\nu_2 = 14$.



Since chi square can take on positive values only, F also can take on only positive values. For most values of ν_1 and ν_2, the distribution of F is generally skewed. Its expected value is given by $E(F) = \nu_2/(\nu_2 - 2)$ and its variance is given by

$$\mathrm{Var}(F) = 2\nu_2^2 \frac{\nu_1 + \nu_2 - 2}{\nu_1(\nu_2 - 2)(\nu_2 - 4)}$$

Since the entire character of the distribution is specified by ν_1 and ν_2, it follows that ν_1 and ν_2 are the parameters of the distribution. The F distribution with $\nu_1 = 7$ and $\nu_2 = 14$ is shown in Figure 14-1. The expected value and variance of this distribution are equal to $E(F) = 1.167$ and $\mathrm{Var}(F) = 8.867$. According to Table A-9, the 95th percentile of the F distribution with $\nu_1 = 7$ and $\nu_2 = 14$ is given by $F_{(7,14)}(.95) = 2.76$, while the 99th percentile is given by $F_{(7,14)}(.99) = 4.28$.

14-2 THE F TEST FOR THE EQUALITY OF TWO VARIANCES FROM NORMAL POPULATIONS

One of the important properties of the F distribution is that it gives rise to an easy-to-perform statistical test that may be used to test whether the variances of two independent *normal* populations are equal in numerical value. The emphasis upon the normality assumption is deliberate since this statistical test is highly dependent upon this assumption. If the variables under study do not have a normal or near-normal distribution, the test to be derived should not be used.

As shown in Section 10-4, $U_1 = (N_1 - 1)S_1^2/\sigma_1^2$ has a chi-square distribution with $\nu_1 = N_1 - 1$ degrees of freedom. In a like manner, $U_2 = (N_2 - 1)S_2^2/\sigma_2^2$ has a chi-square distribution with $\nu_2 = N_2 - 1$ degrees of freedom. According to Theorem 14-1,

$$F = \frac{\nu_2}{\nu_1}\frac{U_1}{U_2} = \frac{N_2 - 1}{N_1 - 1}\frac{(N_1 - 1)S_1^2/\sigma_1^2}{(N_2 - 1)S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2}$$

has an F distribution with $\nu_1 = N_1 - 1$ and $\nu_2 = N_2 - 1$ degrees of freedom. If the hypothesis $H_0\colon \sigma_1^2 = \sigma_2^2$ is true, then the F statistic reduces to

$$F = \frac{S_1^2}{S_2^2}$$

As defined by this equation, F is a variable that can be completely determined from sample values. If, when testing $H_0\colon \sigma_1^2 = \sigma_2^2$, the computed F ratio is too small or too large, then there is good reason to believe that H_0 is false. Thus, an appropriate decision rule for a two-sided test would be to reject H_0 if

$$F < F_{\nu_1,\nu_2}\left(\frac{\alpha}{2}\right) \qquad \text{or if} \qquad F > F_{\nu_1,\nu_2}\left(1 - \frac{\alpha}{2}\right)$$

As might be expected, the F test can always be used in testing one-sided alternatives.

For these cases, all of the probability of rejection should be placed in one tail of the appropriate F distribution.

A rather frequent finding of controlled experimental studies in education is that variability in criterion measures is often affected by the treatment variables. The most common outcome is that the experimental group shows greater variability than the control group or the group that was not given special training. In some studies this finding might be expected on theoretical grounds since education or training frequently tends to increase the variability that exists between individuals or students. Part of this increase in variability is related to interactions between the performance of some subjects and certain experimental treatments. Some individuals react more favorably to special treatments than do others, while others react more negatively than expected. When this increase in variability is suspected or observed, one can always use the F test to determine whether the variability between subjects in an experimental condition when compared to subjects in a control condition is larger than might be expected on the basis of chance.

As an example of this use of the F test, consider the following hypothetical experimental study. In a controlled study on reading in which a number of new teaching methods were being tested, 44 children were randomly assigned to control and experimental conditions. At the end of the study, only 16 of the 22 control subjects remained, while 21 of the 22 experimental subjects had completed the special training. The variances for the two groups of students were as follows: $S_C^2 = 23.6$ and $S_E^2 = 37.9$. Visual inspection of the data suggests that the treatment tended to separate the students. Whether or not this observed sample difference in variability reflects a true parameter difference can be determined by testing $H_0\colon \sigma_C^2 = \sigma_E^2 = \sigma_0^2$ versus the alternative $H_1\colon \sigma_C^2 \neq \sigma_E^2$. As a test statistic, either $F = S_C^2/S_E^2$ or $F = S_E^2/S_C^2$ can be used. For convenience, let $F = S_E^2/S_C^2$ be used as a criterion measure. Since observed outcomes are not permitted to suggest alternative hypotheses, it is imperative that a two-tailed test of hypothesis be conducted. With $\nu_E = 21 - 1 = 20$ and $\nu_C = 16 - 1 = 15$, and with the P(type I error) $= .05$, H_0 should be rejected if $F < F_{(20,15)}(.025) = .389$ or if $F > F_{(20,15)}(.975) = 2.76$. In this case, $F = \frac{37.9}{23.6} = 1.61$. Therefore, there is reason to believe that the variability in the criterion measure is the same for both the experimental and control conditions Note that even if a one-tailed test had been performed, the hypothesis of equal variances would not have been rejected, since the F ratio would have had to exceed 2.33, the 95th percentile of the $F_{(20,15)}$ distribution.

This example illustrates a property of experimental data worthy of further comment. One of the assumptions required for the valid performance of the two-sample t test is that the population variances are equal. In general, this assumption is never exactly true. However, even though it may not be true, moderate departures from a lack of variance equality do not unduly influence the t test. Part of the t test insensitivity or robustness can be traced to the fact that sample variances might be quite discrepant even though the samples are selected from populations in which the variances are equal. In this example, the ratio of the largest sample variance to

the smallest sample variance must exceed 2.33 before the hypothesis of equality at the .05 level would be questioned. Even for moderately sized samples, the ratio of the variances must be quite large before equality of population variances is in doubt. For a one-sided test with H_1: $\sigma_1^2 > \sigma_2^2$, $\alpha = .05$, $N_1 = 25$, and $N_2 = 25$, the ratio of the two variances must exceed 1.98 before one would be willing to conclude that the variances in the population are unequal. This statement can serve as the basis for a simple rule of thumb for deciding when differences in two sample variances suggest inequality in population variances. According to this rule, one would conclude that population variances are unequal if their ratio exceeds 2; otherwise one would behave as though they were equal. In any situation, one can always perform an F test to feel more confident, provided that the variables have a normal or near-normal distribution.

However, some researchers are not willing to perform an F test of equal variances prior to a t test of equal means, since the probability of a type I error for the inferences on means is not equal to the predetermined unconditional α-level of the t test. Performance of the F test changes the exact value of this error so that α is unknown but is most likely larger than expected. As a result, if a researcher believes that the population variances are unequal, then consideration should be given to the Welch–Aspin t^* test as a substitute for sample t tests when making inferences about population centers.

Finally, it should be noted that some researchers would question the use of the F test in our example on the grounds that attrition or subject loss as the experiment proceeded invalidated the study. While it is not always true, it is a fairly safe assumption that the experimental losses are usually found among the unusual elements of the sample, such as sick or truant students. Their loss is almost certain to bias the results. Whether the resulting bias is of major importance no one can say for sure. If a researcher believes that it might be of major importance, then he can always estimate the missing data by substituting the most unfavorable numerical outcomes for the data that have been lost. If, with this conservative set of data, the same decisions are made as with the available set of data, there is little to worry about. If the results should differ, then recourse must be made to the knowledge possessed by the researcher concerning the variables of the study. When relating contradictory results to past knowledge it must be remembered that statistics is only a tool of the behavioral scientist and any unusual results must be viewed with suspicion. As is always the case, good thoughtful judgment and knowledge should be applied—statistics cannot solve all research problems.

14-3 CONFIDENCE INTERVAL FOR THE RATIO OF TWO VARIANCES OF NORMAL POPULATIONS

Just as the t distribution can be used to establish confidence intervals for measures of central tendency, the F distribution can be used to set up confidence intervals for measures of variability. In particular, it can be used to determine confidence intervals for the ratio of two variances. While the confidence intervals for means are valid

even when the parent populations are not normal, provided that the samples are large enough, the same statement does not apply to the confidence intervals for variances. The normality assumption is essential in these cases. Unless one can assume that the parent populations are normal or near normal in form, the confidence-interval statement to be derived should not be used.

Selecting the $\alpha/2$ and $(1-\alpha/2)$ percentiles of the F_{ν_1,ν_2} distribution, we find that

$$P\left[F_{\nu_1,\nu_2}\left(\frac{\alpha}{2}\right) < F < F_{\nu_1,\nu_2}\left(1-\frac{\alpha}{2}\right)\right] = 1-\alpha$$

Thus, a $(1-\alpha)$ percent confidence interval can be obtained from the inequality in the brackets by solving for the ratio of the two variances. Since

$$F = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$$

the inequality statement is equivalent to

$$F_{\nu_1,\nu_2}\left(\frac{\alpha}{2}\right) < \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} < F_{\nu_1,\nu_2}\left(1-\frac{\alpha}{2}\right)$$

Solving these two inequalities for σ_2^2/σ_1^2, we find the final confidence statement given by

$$\frac{S_2^2}{S_1^2}F_{\nu_1,\nu_2}\left(\frac{\alpha}{2}\right) < \frac{\sigma_2^2}{\sigma_1^2} < \frac{S_2^2}{S_1^2}F_{\nu_1,\nu_2}\left(1-\frac{\alpha}{2}\right)$$

If, for the data of Section 14-2, $\sigma_E^2 = \sigma_2^2$ and $\sigma_C^2 = \sigma_1^2$, the 95 percent confidence interval for σ_E^2/σ_C^2 obtained for the observed data is given by

$$\frac{37.9}{23.6}(.389) < \frac{\sigma_E^2}{\sigma_C^2} < \frac{37.9}{23.6}(2.76)$$

or

$$.62 < \frac{\sigma_E^2}{\sigma_C^2} < 4.43$$

In this example, $\sigma_E^2/\sigma_C^2 = 1$ is included in the interval, indicating that $\sigma_E^2 = \sigma_C^2$. This outcome is to be expected since H_0 was not rejected. If H_0 had been rejected, a ratio equal in numerical value to 1 would not be included in the interval. Thus, the confidence interval can be used to test $\sigma_1^2 = \sigma_2^2$. If the interval included 1, H_0 should not be rejected.

14-4 ESTIMATION OF THE VARIANCE FROM K INDEPENDENT SAMPLES

One of the assumptions that will be required in the multiple-sample extension of the t test is that the variances of the individual populations be equal. When the popula-

tion variances are equal, it is possible to obtain a highly efficient estimate of the common value by pooling the data from each of the samples. As shown in Section 11-8,

$$S_p^2 = \frac{(N_1 - 1)\, S_1^2 + (N_2 - 1)\, S_2^2}{N_1 + N_2 - 2}$$

produces an unbiased estimate of σ^2 when independent random samples have been selected from two populations with equal variances. Since the $\mathrm{Var}(S_p^2)$ is considerably less than the $\mathrm{Var}(S_1^2)$ or $\mathrm{Var}(S_2^2)$, it follows that S_p^2 has greater efficiency than either S_1^2 or S_2^2 as an estimate of σ^2. By simple induction, one might suspect that

$$S_p^2 = \frac{(N_1 - 1)\, S_1^2 + (N_2 - 1)\, S_2^2 + \cdots + (N_K - 1)\, S_K^2}{(N_1 - 1) + (N_2 - 1) + \cdots + (N_K - 1)}$$

provides an unbiased estimate of σ^2 that is more efficient than any one-sample estimate. Indeed, this is true, as is now shown in Theorem 14-2 and illustrated in the numerical example included in this discussion.

Theorem 14-2

The pooled estimate of the variance derived from K independent samples selected from populations with equal variances is unbiased.

Proof. By definition, the pooled estimate of the variance is given by the following formula:

$$S_p^2 = \frac{(N_1 - 1)\, S_1^2 + (N_2 - 1)\, S_2^2 + \cdots + (N_K - 1)\, S_K^2}{(N_1 + N_2 + \cdots + N_K) - K}$$

Since $N_1 + N_2 + \cdots + N_K = N$,

$$S_p^2 = \frac{1}{N - K}[(N_1 - 1)\, S_1^2 + (N_2 - 1)\, S_2^2 + \cdots + (N_K - 1) S_K^2]$$

According to Theorem 6-1, the expectation of a sum of random variables is equal to the sum of their expectations Therefore,

$$E(S_p^2) = \frac{1}{N - K}[E\{(N_1 - 1)\, S_1^2\} + E\{(N_2 - 1) S_2^2\} + \cdots + E\{(N_K - 1)\, S_K^2\}]$$

Furthermore,

$$E(S_p^2) = \frac{1}{N - K}[(N_1 - 1)\, E(S_1^2) + (N_2 - 1)\, E(S_2^2) + \cdots + (N_K - 1)\, E(S_K^2)]$$

since according to Theorem 7-4, a constant multiplier may be extracted from the expectation sign. As shown in Section 9-9, sample variances are unbiased estimators of their corresponding population variances. Therefore,

$$E(S_1^2) = \sigma_1^2,\ E(S_2^2) = \sigma_2^2,\ \ldots,\ E(S_K^2) = \sigma_K^2$$

As a result,

$$E(S_p^2) = \frac{1}{N-K}[(N_1 - 1)\sigma_1^2 + (N_2 - 1)\sigma_2^2 + \cdots + (N_K - 1)\sigma_K^2]$$

However, the major assumption of the theorem is that all population variances are equal. Thus

$$E(S_p^2) = \frac{1}{N-K}[(N_1 - 1)\sigma^2 + (N_2 - 1)\sigma^2 + \cdots + (N_K - 1)\sigma^2]$$

and finally,

$$E(S_p^2) = \frac{\sigma^2}{N-K}[(N_1 - 1) + (N_2 - 1) + \cdots + (N_K - 1)]$$

$$= \frac{\sigma^2}{N-K}[N-K]$$

$$= \sigma^2$$

This completes the proof.

As an example of the use of this result, consider the following set of data generated by a study on student risk-taking in answering multiple-choice questions in a statistics examination. The four conditions of the study were as follows:

Condition 1 Students were told that each wrong answer would be scored by subtraction of 2 points.
Condition 2 Students were told that each wrong answer would be scored by subtraction of 1 point.
Condition 3 Students were told that each wrong answer would be scored by subtraction of $\frac{1}{2}$ point.
Condition 4 Students were not given any information concerning the selection of a wrong answer.

When the tests were scored, no points were deducted for wrong answers. Thus, scoring was the same in all four treatment conditions. The scores are as shown in Table 14-1. Thus, for this set of scores the unbiased estimate of the equal population variance is given by

$$S_p = \frac{5(13.37) + 6(43.29) + 5(71.60) + 4(60.70)}{24 - 4} = 46.37$$

14-5 EXPLAINED VARIABILITY

If the possible differences between the experimental conditions in the risk-taking example of the previous section are ignored, it is seen that the total variance of the 24 scores is equal to

$$S_{\text{total}}^2 = \frac{24(12793) - (507)^2}{24(23)} = 90.55$$

Table 14-1. Scores obtained by 24 students on a statistics examination in which students were told that wrong responses would be penalized.

	Treatment 1	*Treatment 2*	*Treatment 3*	*Treatment 4*	*Total*
	8	18	20	37	
	15	17	24	36	
	14	20	35	20	
	17	30	30	22	
	9	14	19	29	
	10	9	40		
		14			
N_k	6	7	6	5	24
$\sum_{i=1}^{N_k} x_{ik}$	73	122	168	144	507
$\sum_{i=1}^{N_k} x_{ik}^2$	955	2,386	5,062	4,390	12,793
$\bar{X}_k$	12.17	17.43	28.00	28.80	21.13
S_k^2	13.37	43.29	71.60	60.70	90.55

Examination of the individual sample variances within a treatment group shows that none are as large as the total variance; all four sample variances are smaller. This suggests that part of the total variance among the 24 scores is related to the variability between the treatments or directions given to the students concerning the scoring of wrong answers. The lowest scores are observed among the students who were told that they would receive the strongest penalties for wrong answers; the top scores are found among the students who would not have been penalized. Thus, it appears that the directions given before the testing have influenced student behavior with respect to the degree to which they will risk guessing on answers they are not sure of.

In a study where the variances of the individual experimental conditions are less than the variance of the total sample, it is said that the experimental conditions account for or explain part of the variability in the criterion variable. Often the degree of accountability is given by a measure of explained variance defined by

$$\hat{\omega}^2 = 1 - \frac{\text{Sum of squares unexplained}}{\text{Sum of squares total}}$$

From this definition, it is seen that the range of $\hat{\omega}^2$ is given by $0 \leqslant \hat{\omega}^2 \leqslant 1$, with 0 representing no effect of the treatment variable and 1 representing complete effect.

For a K-sample model in which it is assumed that the population variances are equal, $\hat{\omega}^2$ reduces to

$$\hat{\omega}^2 = 1 - \frac{(N-K)\,S^2_{\text{pooled}}}{(N-1)\,S^2_{\text{total}}}$$

Thus, for this example,

$$\hat{\omega}^2 = 1 - \frac{(24-4)(46.37)}{(24-1)(90.55)} = 1 - .45 = .55$$

As a percentage measure it is said that the four experimental directions imposed upon the students concerning the consequences of making wrong answers explain about 55 percent of the variability in the final test scores of the total sample. For behavioral variables this is a large percentage and is referred to as a strong treatment effect.

14-6 THE SAMPLING DISTRIBUTION OF S_p^2, THE POOLED ESTIMATE OF σ^2

Suppose it were true that the four samples of the previous section were selected from populations with equal variances. It would then follow that each of the individual variances is an unbiased estimate of the common variance value. As a result, each can be used to set up a confidence interval for σ^2. The 95 percent confidence intervals for the four experimental conditions are as follows:

Treatment	*Lower limit*	*Upper limit*
1	5.21	80.45
2	17.98	209 47
3	27.90	430.80
4	21.80	501.65

These results suggest that σ^2 is between 5.21 and 501.65, a rather uninterestingly broad and useless interval. By Theorem 14-2 it is known that when the population variances are equal, S_p^2 is also an unbiased estimate of σ^2. This estimate of σ^2 is equal to

$$S_p^2 = \frac{5(13.37) + 6(43.29) + 5(71.60) + 4(60.70)}{24-4} = 46.37$$

If the distribution of S_p^2 were known, it would be possible to determine the appropriate confidence interval for σ^2. Fortunately, the distribution of S_p^2 is easy to obtain. As might be expected, its distribution is related to the chi-square distribution.

Theorem 14-3

The distribution of $(N-K)S_p^2/\sigma^2$ is chi square with $(N-K)$ degrees of freedom provided that the underlying variables are normal and the individual random samples are statistically independent.

Proof. By definition,

$$S_p^2 = \frac{1}{N-K}[(N_1-1)S_1^2 + (N_2-1)S_2^2 + \cdots + (N_K-1)S_K^2]$$

As a result,

$$(N-K)S_p^2 = (N_1-1)S_1^2 + (N_2-1)S_2^2 + \cdots + (N_K-1)S_K^2$$

If both sides of this identity are divided by σ^2, it follows that

$$\frac{(N-K)S_p^2}{\sigma^2} = \frac{(N_1-1)S_1^2}{\sigma^2} + \frac{(N_2-1)S_2^2}{\sigma^2} + \cdots + \frac{(N_K-1)S_K^2}{\sigma^2}$$

As is known from Section 10-4, $(N_1-1)S_1^2/\sigma^2$ is distributed as $\chi^2_{N_1-1}$, $(N_2-1)S_2^2/\sigma^2$ is distributed as $\chi^2_{N_2-1}, \ldots,$ and $(N_K-1)S_K^2/\sigma^2$ is distributed as $\chi^2_{N_K-1}$. Therefore,

$$\frac{(N-K)S_p^2}{\sigma^2} = \chi^2_{N_1-1} + \chi^2_{N_2-2} + \cdots + \chi^2_{N_K-1}$$

Since the samples are statistically independent, the variables on the right side of this last equation are statistically independent. Thus, it follows from the additive property of chi square (Theorem 10-7) that

$$\chi^2_{N_1-1} + \chi^2_{N_2-1} + \cdots + \chi^2_{N_K-1} = \chi^2_{N-K}$$

and therefore $(N-K)S_p^2/\sigma^2$ has a chi-square distribution with $\nu = N-K$ degrees of freedom. This completes the proof.

As a result, the $(1-\alpha)$ percent confidence interval for σ^2 is given by

$$\frac{(N-K)S_p^2}{\chi^2_{N-K}(1-\alpha/2)} < \sigma^2 < \frac{(N-K)S_p^2}{\chi^2_{N-K}(\alpha/2)}$$

For the data of Table 14-1,

$$\frac{(24-4)(46.37)}{34.17} < \sigma^2 < \frac{(24-4)(46.37)}{9.59}$$

$$27.14 < \sigma^2 < 96.70$$

Compared to the confidence intervals for each individual sample, this interval is very narrow, reflecting the fact that S_p^2 is a more efficient estimator of σ^2 than any of the individual sample estimators. As was seen earlier in Theorem 14-2, S_p^2 is unbiased as an estimator of σ^2. As a result, it should always be used to estimate σ^2 if it is believed that the variances in the individual populations are equal.

Since S_p^2 is estimated from statistics that measure the variability within each dependent sample, it is often called a within-sample estimate of σ^2 and is frequently noted by MSW or MSR. MSW and MSR are abbreviations for *mean square thin* and *mean square residual*, which will be defined in the sections dealing with analysis of variance and regression. As will be seen, this notation has consider-le utility in analysis-of-variance studies. As a result, one should not hesitate to the "pooled estimate of variance" and the "within estimate of variance" inter-angeably.

7 COCHRAN'S TEST FOR THE EQUALITY OF *K*-POPULATION VARIANCES

e pooled or within estimate of the variance is computed only if it is believed that variances in the parent populations are equal. Often the researcher may not ow for sure whether the population variances are equal; therefore, he may sh to test this hypothesis. A simple test of this hypothesis has been proposed by chran for *equal size samples*. The critical values for this test are presented in ble A-10 of the Appendix. The test statistic for this test is given by

$$C = \frac{\text{largest } S_k^2}{\sum_{k=1}^{K} S_k^2}$$

hypothesis of equal variances is rejected if the computed test statistic exceeds tabled value.

To illustrate the use of this test, consider the data of Table 14-2. These data are en from a 1968 study by Levin and Rohwer. In this part of their study, 48 fifth-de children were randomly assigned to six different experimental conditions in a ial learning experiment. Subjects were required to learn a 14-item serial list in raditional manner, or with the aid of individual phrases, or in the context of a ple sentence. When subjects entered the testing room, they were told by the erimenter to memorize a list of nouns in the order in which they were presented. jects in the ordered-story condition had the nouns logically and meaningfully ied in a single sentence that told a story. Subjects in the ordered-phrase condition the nouns presented in the same order but in phrases. Subjects in the ordered-in condition had the nouns presented in serial order without phrases or sentences. he scrambled condition the nouns appeared in a different order. For the scrambled ry form, the completed list did not convey an understandable story.

The criterion variable of the study was the number of correct responses given the subjects over four test trials. As can be seen, the six variances show some iability. Since each sample is based on $N = 8$ observations, it follows that each nple variance is estimated with $8 - 1 = 7$ degrees of freedom. The value of the statistic is given by $C = \frac{66.70}{317.37} = .21$. To determine whether this represents a nificant difference in population variance, one enters Table A-10 with $K = 6$ $\nu = 7$. For $\alpha = .05$, the decision rule is to reject H_0 if $C > .3980$. Since $.21 < .3980$,

Table 14-2. Number of correct responses given by 48 fifth-grade children to a serial learning task on four test trials.

	Serial list of nouns presented as					
	1 ORDERED STORY	2 ORDERED PHRASES	3 ORDERED NOUNS	4 SCRAMBLED STORY	5 SCRAMBLED PHRASES	6 SCRAMBLED NOUNS
	24	11	30	28	21	14
	23	17	14	19	7	19
	43	8	21	30	24	19
	39	25	25	15	13	35
	35	19	23	11	31	18
	24	21	31	20	18	6
	30	10	11	17	11	18
	28	22	19	13	20	14
$\sum_{i=1}^{8} X_{ik}$	246	133	174	153	145	143
$\sum_{i=1}^{8} X_{ik}^2$	7960	2485	4134	3249	3041	3023
$\bar{X}_k$	30 75	16.62	21 75	19 13	18 13	17 88
S_k^2	56.50	39.13	49 93	46.13	58 98	66 70

one would not reject the hypothesis of equal variances. It should be emphasized that Cochran's test is for equal-sized samples. To avoid the problem of unequal sample size, one can always randomly discard data to make equal-sized samples for this test. Of course, this would only be done if the sample sizes are approximately equal. When the sample sizes are quite unequal, other procedures are available. The most popular test is Bartlett's test for equal variances, which is discussed in Exercise 14-7.

Note that the sample variances of this example appear to be quite variable, and yet by Cochran's test they must be considered as being equal in the populations. This finding is of applied interest because the *K*-sample extension of the *t* test requires equality of variance for its valid use. As this example suggests, a researcher need not be too concerned if the sample variances differ among themselves to a moderate degree. Such variability is to be expected. If, however, a researcher thinks that the observed variability in sample variances exceeds chance expectations, he can always test his hunch by means of Cochran's test, provided that the sample sizes are equal.

14-8 THE RELATIONSHIPS AMONG THE t, F, AND χ^2 DISTRIBUTIONS

As was seen in Section 10-4,

$$U_1 = Z^2 = \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{N}}\right)^2$$

Table 14-3. Confidence intervals and hypothesis-testing model for equality of variances.

Case	*Hypothesis*	*Test statistic*	*Confidence interval*	*Assumptions*
10	$H_0 .\ \sigma_1^2 = \sigma_2^2 = \sigma_0^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$ $\nu_1 = (N_1 - 1),\ \nu_2 = (N_2 - 1)$	$F_{\nu_1,\nu_2}\left(\frac{\alpha}{2}\right)\frac{S_2^2}{S_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < F_{\nu_1,\nu_2}\left(1 - \frac{\alpha}{2}\right)\frac{S_2^2}{S_1^2}$	1. Independence between samples 2. Independence within samples 3. Normality

ν_1 ... es of freedom and

$\nu_2 =$ ... grees of freedom While it is extremely
$_1$ ar ... statistically independent. Therefore,

$= \Big(\quad = t_{\nu_2}^2$

ed a ... follows that $F_{1,N-1} = t_{N-1}^2$. Thus, when
nd F ... ely allied to one another. Examination
at th ... tionships are indeed true. For example,
5) = ... and $t_{18}(.975) = 2.10$. In Table A-9 it is
$\sqrt{4}$ ... 10, it is seen that the relationship is

thou ... then U_2/ν_2 tends to approach unity.
ed si ... $U_2) = \nu_2$. Therefore,

ar(U ... ν_2, it follows that

$= \frac{2}{\nu_2}$

ts, ... ν_2) tends to 0 so that U_2/ν_2 tends to its
hat

0, a ... ample of this convergence, Tables A-7
2 are ... at $15F_{15,\infty}(.95) = 15(1.67) = 25.0$. Thus,
zed ... $F_{\nu_1,\nu_2}(\alpha)$ is quite close in form to $\chi_{\nu_1}^2(\alpha)$.
... orems 14-4 and 14-5, which are stated

Theorem 14-4

The t distribution with ν_2 degrees of freedom is associated to the F distribution with $\nu_1 = 1$ and ν_2 degrees of freedom by the equation $t^2_{\nu_2} = F_{1,\nu_2}$.

Theorem 14-5

The χ^2 distribution with ν_1 degrees of freedom is associated to the F distribution with ν_1 and $\nu_2 = \infty$ degrees of freedom by the equation $\chi^2_{\nu_1} = \nu_1 F_{\nu_1,\infty}$.

14-9 TESTS AND CONFIDENCE INTERVALS FOR VARIANCES IN THE TWO- AND K-SAMPLE MODEL

In Table 14-3, a summary of the confidence interval and hypothesis-testing model for equality of two or more variances is presented. For the Cochran model, an example was presented in which the hypothesis H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2 = \sigma_0^2$ was not rejected. When this occurs, one usually places a confidence interval about σ_0^2, based upon the pooled estimate of the variances and the chi-square distribution with $\nu = N - K$. If the hypothesis were rejected, then one has a problem because a possible reason for the rejection is unknown. One way to identify possible reasons for rejection is to compute all possible two-variance confidence intervals based upon the appropriate F distributions. Since each interval has an associated probability of a type I error, it is customary to control the overall probability of at least one type I error. This control is achieved in the following manner. If there are K samples, the number of two-sample confidence intervals is given by $Q = \binom{K}{2}$. If the entire set of intervals is to be controlled with P(at least one type I error) $= \alpha$, confidence intervals should be determined at $\alpha_0 = \alpha/Q$. For the data of Table 14-2, $K = 6$. If the number of confidence intervals is to be controlled at $\alpha = .05$, then each confidence interval should be controlled at $\alpha_0 = \frac{.05}{15} = .0033$. This will guarantee that the probability of at least one type I error in the set of 15 intervals is $\leqslant .05$.

14-10 SUMMARY

In this chapter, the important F distribution was introduced to produce a test of the hypothesis H_0: $\sigma_1^2 = \sigma_2^2$ versus H_1: $\sigma_1^2 \neq \sigma_2^2$, under the assumption that two independent random samples had been selected from normally distributed universes. As will be seen in the following chapter, this same distribution can be used to test other more interesting statistical hypotheses.

As an example of this use of the test of $\sigma_1^2 = \sigma_2^2$, consider a hypothetical study in which 16 college sophomores are told to learn a list of 10 pairs of nonsense words in as few trials as possible, while 16 other students are told to learn the list in 20 trials, for if they do not, they will be given an electric shock after each trial until they can give the list from memory. In addition to determining whether the mean number of trials is influenced by the experimental conditions, one might wonder what the effects would be upon the variability in number of trials it takes to learn the task.

Table 14-3. Confidence intervals and hypothesis-testing model for equality of variances.

Case	*Hypothesis*	*Test statistic*	*Confidence interval*	*Assumptions*
10	$H_0 . \sigma_1^2 = \sigma_2^2 = \sigma_0^2$ $H_1 : \sigma_1^2 \neq \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$ $\nu_1 = (N_1 - 1),\ \nu_2 = (N_2 - 1)$	$F_{\nu 1,\nu 2}\left(\frac{\alpha}{2}\right)\frac{S_2^2}{S_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < F_{\nu 1,\nu 2}\left(1 - \frac{\alpha}{2}\right)\frac{S_2^2}{S_1^2}$	1. Independence between samples 2 Independence within samples 3 Normality
11	$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2 = \sigma_0^2$ $H_1 : H_0$ is false	$C = \frac{\text{largest } S_k^2}{\sum_{k=1}^{K} S_k^2}$	$F_{\nu k_1,\nu k_2}\left(\frac{\alpha_0}{2}\right)\frac{S_{k2}^2}{S_{k1}^2} < \frac{\sigma_{k2}^2}{\sigma_{k1}^2}$ $< F_{\nu k_1,\nu k_2}\left(1 - \frac{\alpha_0}{2}\right)\frac{S_{k2}^2}{S_{k1}^2}$ $\alpha_0 = \frac{\alpha}{\binom{K}{2}}$ $k_1, k_2 = 1, 2, \ldots, K$, but $k_1 < k_2$	1 Independence between samples 2. Independence within samples 3. Normality
12	$H_0 . \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$\chi^2 = \frac{(N - K)S_p^2}{\sigma_0^2}$ where $\nu = (N - K)$	$\frac{(N - K)S_p^2}{\chi^2_{N-K}(1 - \alpha/2)} < \sigma^2 < \frac{(N - K)S_p^2}{\chi^2_{N-K}(\alpha/2)}$	1. Independence between samples 2. Independence within samples 3. Normality

It might be hypothesized that students in the experimental condition will show greater variance because of induced anxiety. Thus, the hypothesis to test is $H_0\colon \sigma_C^2 = \sigma_E^2$ versus $H_1\colon \sigma_C^2 < \sigma_E^2$. For this test one could use $F = S_E^2/S_C^2$ and reject H_0 at $\alpha = .05$, if $F > F_{15,15}(.95) = 2.40$.

Suppose that the study had been conducted and that $S_E^2 = 17.3$ and $S_C^2 = 12.2$. For these data, $F = \frac{17.3}{12.2} = 1.58$, which does not lead to rejection of the hypothesis. Thus, even though the difference in variance is in the predicted direction, the experimental hypothesis cannot be defended. If the hypothesis had been rejected, a one-sided confidence interval of the form

$$\frac{\sigma_E^2}{\sigma_C^2} > \frac{S_E^2}{S_C^2} F_{\nu_C,\nu_E}(\alpha)$$

could have been determined to see how large the ratio of the population variance might be.

When the number of observations in K different samples are equal, then one can test the hypothesis $H_0\colon \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2 = \sigma_0^2$ by means of the Cochran statistic:

$$C = \frac{S_{\text{largest}}^2}{\sum_{k=1}^{K} S_k^2}$$

If this hypothesis is not rejected, then one can estimate the common value of σ_0^2 by computing

$$S_p^2 = \sum_{k=1}^{K} \frac{S^2}{K}$$

This result is a special case of the general form that is valid even when the sample sizes are unequal, provided that it is assumed that the variances of the K populations are equal. When these conditions are satisfied, an unbiased estimate of σ_0^2 is given by

$$S_p^2 = \frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2 + \cdots + (N_K - 1)S_K^2}{(N_1 + N_2 + \cdots + N_K) - K} = \sum_{k=1}^{K} \frac{(N_k - 1)S_k^2}{N - K}$$

This estimate has a distribution related to the χ^2 distribution with $\nu = N_1 + N_2 + \cdots + N_K - K = N - K$, so that a $(1 - \alpha)$ percent confidence interval for σ_0^2 is given by

$$\frac{(N-K)S_p^2}{\chi_\nu^2\left(1 - \frac{\alpha}{2}\right)} < \sigma_0^2 < \frac{(N-K)S_p^2}{\chi_\nu^2\left(\frac{\alpha}{2}\right)}$$

Finally, it was noted that the F distribution was intimately connected with the χ^2 and t distributions. In fact, F_{ν_1,ν_2} was defined as the ratio of two independent chi-square variables, each divided by its degrees of freedom. Symbolically,

$$F_{\nu_1,\nu_2} = \frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}$$

If ν_2 tends to infinity, F_{ν_1,ν_2} tends to $(1/\nu_1)\chi^2_{\nu_1}$, so that in the limit, $\chi^2_{\nu_1} = \nu_1 F_{\nu_1,\infty}$. Also, if $\nu_1 = 1$, then $F_{1,\nu_2} = t^2_{\nu_2}$. The importance of these relationships will be noted in the following chapters.

EXERCISES

***14-1.** In Table A-9, find

(*a*) $F_{5,18}(.005)$	(*e*) $F_{21,53}(.90)$
(*b*) $F_{18,5}(.01)$	(*f*) $F_{60,120}(.95)$
(*c*) $F_{9,9}(.025)$	(*g*) $F_{40,\infty}(.975)$
(*d*) $F_{20,30}(.05)$	(*h*) $t^2_{20}(.99)$

***14-2.** Find the $E(F)$ for each of the F distributions of Exercise 14-1. What does this tell you about the distribution of F, if $\sigma_1^2 = \sigma_2^2$?

14-3. In a school with a large number of culturally deprived children, it was found among 18 minority students that the standard deviation in IQ units as measured by the Kuhlman–Anderson intelligence test was given by $S_m = 12.6$, while for a corresponding set of 23 nonminority students taken from the same classes, $S_{\bar{m}} = 13.9$. Does this evidence support the hypothesis that the variability in IQ scores is the same for these two populations? Exactly what are the two populations of the study?

14-4. As part of a medical examination in which heights of entering male and female freshmen were measured, it was found that for 193 males, $S_M = 2.9$ inches and for 118 females, $S_F = 1.9$ inches. Does this indicate that males are more variable in height than females? What have you assumed in reaching your decision? Do you think the assumptions are satisfied? Why?

***14-5.** Are the variances of the two universes of Exercise 8-5 equal or different?

***14-6.** When the sample sizes are unequal, one can still perform the Cochran test for equal variances by randomly discarding data to make all sample sizes equal. With this in mind, randomly remove data from Table 14-1 to produce four samples of size 5. On this reduced sample size, test the hypothesis of equal variances for the four conditions of the study. If the hypothesis is rejected, use the method of Section 14-10. to locate possible sources for the rejection.

***14-7.** When sample sizes are large and unequal, one can perform the Bartlett test for equality of variances. The test statistic for this test is given by

$$\chi^2 = \frac{2.303}{C}\left[(N-K)\log_{10} S_p^2 - \sum_{k=1}^{K}(N_k - 1)\log_{10} S_k^2\right]$$

where

$$C = 1 + \frac{1}{3(K-1)}\left(\sum_{k=1}^{K}\frac{1}{N_k - 1} - \frac{1}{N-K}\right)$$

which, when H_0 is true, is approximately χ^2 with $\nu = K - 1$. Using the data of Exercise 8-7, test the hypothesis of equal variances in yearly incomes for the five sets of data defined by the number of completed semesters of junior college. What does the resulting test suggest to you about the population variances? If H_0 is rejected, use the method of Section 14-10 to locate possible sources for the rejection.

***14-8.** Set up the 97.5 percent confidence intervals for σ_1^2/σ_0^2 and σ_5^2/σ_0^2 of Exercise 11-2. What do the resulting intervals suggest to you about the hypothesis $H_0: \sigma_0^2 = \sigma_1^2 = \sigma_5^2$? Can you give a reason why two confidence intervals with $\alpha = .025$ are better than two confidence intervals with $\alpha = .05$ for answering this question?

***14-9.** Set up the 95 percent confidence interval for the ratio of the variances of Exercise 11-7. What must you assume in interpreting this interval? Is the interpretation valid in this case? Why?

***14-10.** Answer Exercise 11-9*c*, *d*, and *e* in terms of the methods presented in this chapter.

15 HYPOTHESIS TESTING IN THE *K*-SAMPLE QUANTITATIVE ANALYSIS-OF-VARIANCE MODEL

THE PROOF OF PARKINSON

In 1955, British Historian C. Northcote Parkinson puckishly formulated the basic law of bureaucracy that bears his name: work expands to fit the time at hand for doing it. Parkinson himself regarded his "law" as satire; inevitably, several American psychologists have decided to take it seriously.. .

One of these investigators is Social Psychologist Elliot Aronson of the University of Texas, who became interested in the law after suffering through a Parkinsonian procrastination of his own making · he took three desultory summer weeks to prepare a lecture that could have been written in three hours. Deciding to test the work-delaying proclivities of others, he divided a number of volunteer students into two groups. Those in one section were allowed five minutes to prepare a talk on the subject of smoking, the others were given 15 minutes for the job. Aronson then gave each group a new but similar chore, allowing them to take as much time as they wanted. The five-minute students managed to finish the job in accordance with their original deadline; the others, having initially decided that the job required more time, took an average of eight minutes to complete the assignment

15-1 INTRODUCTORY REMARKS CONCERNING THE K-SAMPLE PROBLEM

The "classical" two-sample tests presented in Chapter 13 are perhaps the most frequently employed statistical tests of behavioral researchers. Close behind them in relative frequency of use are the multiple-sample analogs of these tests. At first thought it might be assumed that the multiple K-sample analogs to the two-sample tests simply consist of all possible two-sample tests that could be generated from K groups of data. Unfortunately, this is not the case. Instead, omnibus tests based on a test statistic related to the F or χ^2 distributions are used in performing the corresponding tests. In this chapter, tests based on the F distribution are presented. Tests utilizing the χ^2 distribution are discussed in Chapter 16.

15-2 THE ONE-WAY ANALYSIS OF VARIANCE—THE K-SAMPLE EXTENSION OF THE TWO-SAMPLE t TEST

The main elements of the K-sample problem are easily demonstrated by reviewing the data presented in Table 14-1. The relationships that exist across the various conditions of the study are quite typical of what is generally observed in behavioral research. Comparison of the variances of the individual samples to the variance based on all observations, independent of sample membership, shows that the variances within the individual samples ($S_1^2 = 13.37$, $S_2^2 = 43.29$, $S_3^2 = 71.60$, $S_4^2 = 60.70$) are considerably less than the variance of the total sample ($S^2 = 90.55$). As was indicated, these lower values for the individual sample variances indicate that part of the total variance is related to the differences that exist between groups as a result of the treatment condition imposed upon the subjects. As was noted, the manipulated variable accounted for $\hat{\omega}^2 = 55$ percent of the total variability. When the explained variance is this large, it follows that the centers of the distributions defined by the various treatments are not equal to the same numerical value. This supposition corresponds to a denial of the statistical hypothesis $H_0\colon \mu_1 = \mu_2 = \cdots = \mu_K = \mu_0$, which states that all distribution centers are equal in numerical value. Surprisingly, this statistical hypothesis about population centers can be evaluated by means of a test statistic that is used to answer statistical questions about population variances and that has an F distribution when the population centers are equal in numerical value.

To help understand the rationale of this statistical test, reconsider the data of Table 14-1, but assume that the sample sizes are all equal to $N_0 = 6$. Later this assumption is dropped and the derivation of this procedure is provided for unequal sample sizes. For the data of Table 14-1, $\bar{X}_1 = 12.17$, $\bar{X}_2 = 17.43$, $\bar{X}_3 = 28.00$, $\bar{X}_4 = 28.80$, and $\bar{X} = 21.13$. If, in addition, it is assumed that $N_1 = N_2 = N_3 = N_4 = 6$, and that

1. Independent random samples are selected from each population defined by the experimental conditions
2. The criterion variable of the study is normally distributed
3. The population means are equal so that $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_0$
4. The population variances are equal so that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_0^2$

then it follows that the four distributions are exactly identical since all are $N(\mu_0,\sigma_0^2)$. This is illustrated in Figure 15-1 as H_0: the model under test. When the distributions are identical it follows that the four sample means are really a random sample of means drawn from the sampling distribution of means generated from $N(\mu_0,\sigma_0^2)$. Since it is assumed that each sample is based on six observations, it follows that the sampling distribution of the means is $N(\mu_0,\sigma_0^2/6)$.

If the variance of the observed sample means is now computed, it follows that

$$S_{\bar{X}}^2 = \frac{1}{K-1}\sum_{k=1}^{K}(\bar{X}_K - \bar{X})^2$$

$$= \frac{(12.17-21.13)^2 + (17.43-21.13)^2 + (28.00-21.13)^2 + (28.80-21.13)^2}{4-1}$$

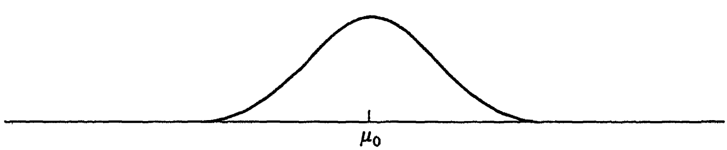
$$= 64.156$$

Under the assumptions, this is an unbiased estimate of $\sigma_{\bar{X}}^2 = \sigma_0^2/6$. Since $6\sigma_{\bar{X}}^2 = \sigma_0^2$, it follows that $6S_{\bar{X}}^2$ is an estimate of σ_0^2, which in this case is equal to 6(64.156) = 384.94. This estimate of σ_0^2 is based on the variability between the means and is called the between sample estimate of σ_0^2. It is frequently denoted as the *mean square between groups* and abbreviated as MSB. Since it is based on four sample means, it is really an estimate of σ_0^2 based on $\nu_1 = K-1 = 4-1 = 3$ degrees of freedom.

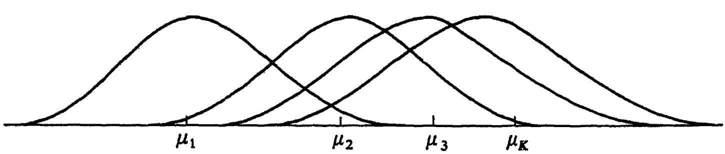
It was also seen in Section 14-4 that the within sample estimate of σ_0^2, based on $\nu_2 = N-K = 24-4 = 20$ degrees of freedom, was given by $S_p^2 = 46.37$. Since this estimate of σ_0^2 is based on the variability within the samples, it is called the *mean square within groups* and abbreviated as MSW. Thus, it is possible to compute two

Figure 15-1. The analysis-of-variance model under H_0 and H_1

H_0: The model under test: $\mu_1 = \mu_2 = \cdots = \mu_K = \mu_0$



H_1: The alternative model: At least two μ's are different



different estimates of σ_0^2 from the same set of data, one based on the variability between the samples and one based on the variability within the samples. In addition, they can be tested for equality by the F test. In this case,

$$F = \frac{\text{MSB}}{\text{MSW}} = \frac{384.94}{46.37} = 8.30$$

Since $F_{3,20}(.95) = 3.10$, it follows that the two estimates of σ_0^2 are statistically different. This implies that the assumptions under which the test was made are suspect. Of the four assumptions, the one involving equality of means is most suspect since the deviations of the individual means from the grand mean are quite large. In this case, the values of the deviations are (–8.96, –3.70, 6.87, 7.67). For this reason, the hypothesis of equal expected values should be rejected.

For completeness, a derivation of this test is provided for the special case $N_1 = N_2 = \cdots = N_K = N_0$. However, the theorem is stated for unequal sample sizes.

Theorem 15-1

If K independent random samples are selected from normally distributed universes in which the variances are equal, then the test statistic for testing the hypothesis H_0: $\mu_1 = \mu_2 = \cdots = \mu_K = \mu_0$ versus the alternative $H_1 \cdot H_0$ is false is given by

$$F = \frac{\text{mean square between}}{\text{mean square within}} = \frac{\text{MSB}}{\text{MSW}}$$

$$= \frac{1/(K-1) \sum_{k=1}^{K} N_k(\bar{X}_k - \bar{X})^2}{1/(N-K) \sum_{k=1}^{K} (N_k - 1)\, S_k^2}$$

which, when H_0 is true, has an F distribution with parameters $\nu_1 = K - 1$ and $\nu_2 = N - K$, and where

N_k = size of the kth sample
$\bar{X}_k$ = mean of the kth sample
S_k^2 = variance of the kth sample
$\bar{X}$ = weighted average of the sample means
$N = N_1 + N_2 + \cdots + N_K$

Proof. As a first step in deriving the test statistic, assume

1. $\mu_1 = \mu_2 = \cdots = \mu_K = \mu_0$
2. $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2 = \sigma_0^2$
3. $N_1 = N_2 = \cdots = N_K = N_0$
4. All populations are normal in form

Under these conditions it follows that $\bar{X}_1$, the mean of the first sample, is a random observation from a theoretical sampling distribution that is $N(\mu_0,\sigma_0^2/N_0)$. In a like manner, $\bar{X}_2$ is a random observation from a distribution that is $N(\mu_0,\sigma_0^2/N_0)$, as is $\bar{X}_3, \bar{X}_4, \ldots, \bar{X}_K$. Thus, it follows that $(\bar{X}_1,\bar{X}_2,\ldots,\bar{X}_K)$ is in reality a random sample of K observations from a theoretical sampling distribution of means for which

$$E(\bar{X}_k) = \mu_0 \qquad \text{and} \qquad \operatorname{Var}(\bar{X}_k) = \frac{\sigma_0^2}{N_0}$$

Since it is assumed that the sample sizes are equal, the mean of the sample means is given by

$$\bar{X} = \frac{1}{K}\sum_{k=1}^{K} \bar{X}_k$$

Thus

$$S_{\bar{X}}^2 = \frac{1}{K-1}\sum_{k=1}^{K} (\bar{X}_k - \bar{X})^2$$

must be an unbiased estimate of $\sigma_{\bar{X}}^2 = \sigma_0^2/N_0$. That is,

$$E(S_{\bar{X}}^2) = E\left[\frac{1}{K-1}\sum_{k=1}^{K} (\bar{X}_k - \bar{X})^2\right] = \frac{\sigma_0^2}{N_0}$$

and as a result,

$$E(N_0 S_{\bar{X}}^2) = E\left[\frac{N_0}{K-1}\sum_{k=1}^{K} (\bar{X}_k - \bar{X})^2\right] = \sigma_0^2$$

From this last statement, it is seen that the algebraic expression within the brackets is an unbiased estimator of σ_0^2. This estimate is based upon the variability that exists between the sample averages and it is called the between sample estimate of the variance. Sometimes it is referred to as the *mean square between groups*. In this book, this estimate will be denoted MSB.

Since MSB is an estimate of σ_0^2 based upon K independent observations, it follows that $U_1 = (K-1)\text{MSB}/\sigma_0^2$ has a chi-square distribution with $\nu_1 = (K-1)$ degrees of freedom. This last result follows from either of two arguments. If the parent populations are normal, then the sampling distribution of means is normal. If the parent populations are not normal, then the central limit theorem must be invoked to ensure that the distribution of the means is close to normal in form. For most of the distributions encountered in behavioral research, the convergence of sample means to a normal form is quite rapid for relatively small sample sizes and so exact normality of the underlying distribution is not absolutely necessary.

In Section 14-4, S_p^2 was derived as an estimate of σ_0^2 based on the variability that exists within the samples. This estimate of σ^2 has been denoted by MSW. In addition, it should be noted that S_p^2 also is an estimate of σ^2. Thus, MSW and S_p^2 can be interchanged depending upon the context.

Also, as shown in Section 14-6, $U_2 = (N-K)\text{MSW}/\sigma_0^2$ has a chi-square distribution with $\nu_2 = (N-K)$ degrees of freedom when the parent populations are normal in form.

It can be shown by means of advanced methods that MSB and MSW are statistically independent estimates of σ_0^2 if H_0 is true. Therefore, an F test can be created to test H_0: $E(\text{MSB}) = E(\text{MSW}) = \sigma_0^2$. Under the assumed conditions and unequal sample sizes,

$$F = \frac{U_1/\nu_1}{U_2/\nu_2} = \frac{\dfrac{(K-1)\,\text{MSB}/\sigma_0^2}{(K-1)}}{\dfrac{(N-K)\text{MSW}/\sigma_0^2}{N-K}} = \frac{\text{MSB}}{\text{MSW}} = \frac{1(K-1)\sum_{k=1}^{K} N_k(\bar{X}_k - \bar{X})^2}{1(N-K)\sum_{k=1}^{K}(N_k-1)S_k^2}$$

For emphasis, it is repeated that MSB is an unbiased estimate of σ_0^2 only if H_0 is true. If H_0 is false, this estimate is biased and is too large an estimate of the common variance. While the proof is beyond the methods of this book, it is possible to show that if H_0 is true, then

$$E(\text{MSB}|H_0) = \sigma_0^2$$

but if H_1 is true, then for unequal sample sizes,

$$E(\text{MSB}|H_1) = \sigma_0^2 + \frac{1}{K-1}\sum_{k=1}^{K} N_k(\mu_k - \mu_0)^2$$

Thus, MSB is a good index of the truthfulness of H_0: when it is small, equality of centers is a reasonable decision, but when it is large it is more reasonable to conclude that centers are different.

When H_0 is true, F has an F distribution with $\nu_1 = (K-1)$ and $\nu_2 = (N-K)$ degrees of freedom since $E(\text{MSB}|H_0) = E(\text{MSW}|H_0)$. If H_0 is false, MSB will tend to be larger than MSW since $E(\text{MSB}|H_1) > E(\text{MSW}|H_1)$. Therefore, a large F ratio is compatible with the truth of H_1, the alternative hypothesis. This suggests that H_0 should be rejected if $F > F_{K-1,N-K}(1-\alpha)$.

If H_0 is false, then at least one of the population means must be different from the others. When this is true, the variability between the sample means will be large. Thus, if H_0 is not rejected, there is reason to believe that $\mu_1 = \mu_2 = \cdots = \mu_K$, but if H_0 is rejected then the truth of this statement is doubtful. Thus, the test $H_0 \cdot E(\text{MSB}) = E(\text{MSW}) = \sigma_0^2$ is equivalent to the test of $H_0 .\ \mu_1 = \mu_2 = \cdots = \mu_K = \mu_0$ versus H_1: H_0 is false. This indicates that this particular test of equality of two independent estimates of the same variance is in reality a test of equality of means. This completes the proof.

In summary, the F test, in the form presented here, is used to test the hypothesis that K independent populations have equal expected values. The assumptions required for this test are:

1. Each parent population is normal in form or the sample sizes are large enough to ensure that the distribution of means is near normal in form.
2. The samples from each population are random samples.
3. The samples from each population are selected independently from each of the other samples.
4. The variances in each of the populations are equal.

When discussing the F test for the equality of two variances, it was seen that sample variances could be quite discrepant and yet be nonstatistically different from one another. This suggests that assumption 4 can be relaxed to a moderate degree. The same is true for the normality assumption. The F test is quite robust to moderate departures from these two assumptions. There is some evidence that the variances can be quite different, provided that the sample sizes are equal. When the sample sizes are equal and the population variances are unequal, the actual probability of a type I error is very close to the nominal probability of a type I error. However, this may not be true for the probability of a type II error. Sometimes it is larger and sometimes it is smaller. Unfortunately, it is not possible to give any sound guidelines that a researcher can follow when the population variances are different. However, if the population variances are different and if the sample sizes are quite variable, then perhaps the best advice is not to perform the F test but to do the test discussed in Section 15-7.

15-3 COMPUTATION FORMULAS FOR THE SUM OF SQUARES IN AN ANALYSIS-OF-VARIANCE MODEL

Sometimes the computations required for the analysis of variance can be simplified if one bysteps the determination of the sample means and sample variances. While it is not recommended that this step be ignored, one can still use the computing formulas to simplify the computations needed to obtain F. For this simplification, one first computes the $T_{.k} = X_{1k} + X_{2k} + \cdots + X_{N_k k} =$ total for the kth sample. Then the grand total of all the observations is given by $T_{..} = T_{.1} + T_{.2} + \cdots + T_{.K}$, and

$$\bar{X}_1 = \frac{T_{.1}}{N_1},\ \bar{X}_2 = \frac{T_{.2}}{N_2},\ \ldots,\ \bar{X}_K = \frac{T_{.K}}{N_K} \qquad \text{and} \qquad \bar{X} = \frac{T_{..}}{N}$$

Furthermore,

1. The sum of squares between groups or samples is given by

$$\text{SSB} = \sum_{k=1}^{K} N_k(\bar{X}_k - \bar{X})^2 = \sum_{k=1}^{K} N_k(\bar{X}_k^2 - 2\bar{X}\bar{X}_k + \bar{X}^2)$$

$$= \sum_{k=1}^{K} N_k \bar{X}_k^2 - 2\bar{X} \sum_{k=1}^{K} N_k \bar{X}_k + \bar{X}^2 \sum_{k=1}^{K} N_k$$

$$= \sum_{k=1}^{K} N_k \left(\frac{T_{.k}}{N_k}\right)^2 - 2(NT_{..})T_{..} + \left(\frac{T_{..}}{N}\right)^2 N$$

$$= \sum_{k=1}^{K} \frac{T_{.k}^2}{N_k} - \frac{T_{..}^2}{N}$$

2. The sum of squares within groups or samples is given by

$$\text{SSW} = \sum_{k=1}^{K} (N_k - 1) S_k^2 = \sum_{k=1}^{K} (N_k - 1) \left(\frac{N_k \sum_{i=1}^{N_k} X_{ik}^2 - \left(\sum_{i=1}^{N_k} X_{ik} \right)^2}{N_k(N_k - 1)} \right)$$

$$= \sum_{k=1}^{K} \left(\sum_{i=1}^{N_k} X_{ik}^2 - \frac{1}{N_k} \left(\sum_{i=1}^{N_k} X_{ik} \right)^2 \right)$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{N_k} X_{ik}^2 - \sum_{k=1}^{K} \frac{1}{N_k} (T_{.k})^2$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{N_k} X_{ik}^2 - \sum_{k=1}^{K} \frac{T_{.k}^2}{N_k}$$

3. The sum of squares total is given by

$$\text{SST} = \sum_{k=1}^{K} \sum_{i=1}^{N_k} (X_{ik} - \bar{X})^2 = \sum_{k=1}^{K} \sum_{i=1}^{N_k} (X_{ik}^2 - 2\bar{X}X_{ik} + \bar{X}^2)$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{N_k} X_{ik}^2 - 2\bar{X} \sum_{k=1}^{K} \sum_{i=1}^{N_k} X_{ik} + \bar{X}^2 \sum_{k=1}^{K} \sum_{i=1}^{N_k} 1$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{N_k} X_{ik}^2 - 2\left(\frac{T_{..}}{N}\right)(NT_{..}) + \left(\frac{T}{N}\right)^2 N$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{N_k} X_{ik}^2 - \frac{T_{..}^2}{N}$$

If

$$\text{I} = \sum_{k=1}^{K} \sum_{i=1}^{N_k} X_{ik}^2$$

$$\text{II} = \frac{T_{..}^2}{N}$$

and

$$\text{III}_C = \sum_{k=1}^{K} \frac{T_{.k}^2}{N_k}$$

then

1. $\text{SSB} = \text{III}_C - \text{II}$
2. $\text{SSW} = \text{I} - \text{III}_C$
3. $\text{SST} = \text{I} - \text{II}$

For the data of Table 14-1,

1. $\mathrm{I} = 12793$

2. $\mathrm{II} = \dfrac{(507)^2}{24} = 10710.38$

3. $\mathrm{III}_C = \dfrac{(73)^2}{6} + \dfrac{(122)^2}{7} + \dfrac{(168)^2}{6} + \dfrac{(144)^2}{5} = 11865.66$

4. $\mathrm{SSB} = 11865.66 - 10710.38 = 1155.28$

5. $\mathrm{SSW} = 12793 \quad - 11865.66 = \ \ 957.34$

6. $\mathrm{SST} = 12793 \quad - 10710.38 = 2082.62$

15-4 THE ANALYSIS-OF-VARIANCE TABLE

In journal articles and research reports it is customary to summarize the findings of a study in what is called an analysis-of-variance table. The analysis-of-variance table of the data of Table 14-1 is as shown in Table 15-1. The figures and column headings of this table are self-explanatory, having been discussed in great detail in previous sections.

Table 15-1. The analysis-of-variance table of the data of Table 14-1.

Source of variation	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F-value*	*Expected mean square*
Between treatments	3	1155.28	385 09	8 30	$\sigma^2 + \dfrac{1}{K-1}\sum N_k(\mu_k - \mu)^2$
Within treatments	20	927.34	46 37		σ^2
Total	23	2082.62	90.55		

15-5 SCHEFFÉ'S METHOD OF MULTIPLE CONTRASTS—*POST HOC* COMPARISONS

While the rejection of the hypothesis of equal treatment effects may be statistically interesting, it is, in general, not very useful. To know that the treatments differ in their effects upon the criterion variable does not say very much, because one still does not know which treatments differ from one another nor to what degree they differ. The performance of all possible two-treatment t tests might help locate the source of variance, if one uses the proper significance levels and tabled t values. However, the employment of multiple t tests entails possible increases in the risk

of making a type I error and for that reason it is not usually recommended Furthermore, since the proper multiple t test procedure is so dependent upon the particular study being analyzed, it is difficult to present an adequate discussion of appropriate methods in a beginning statistics book. Therefore, an omnibus method of analysis that has gained wide acceptance will be presented and illustrated. This method was derived by Scheffé and is referred to as Scheffé's method of multiple contrasts.

In terms of a one-way analysis-of-variance design, *a contrast* is defined as a weighted sum of the expected values of the K populations such that the coefficients that serve as multipliers on the expected values have their sum equal to 0. The most typical example of a contrast is given by $\psi = a_1\mu_1 + a_2\mu_2 + \cdots + a_K\mu_K$, where $a_1 + a_2 + \cdots + a_K = 0$. The most familiar example of a contrast is given by $\psi = \mu_1 - \mu_2$ with $a_1 + a_2 = 1 - 1 = 0$. Contrasts of this form are used to compare two expected values. They were studied in great detail in Chapters 11 and 13. The basic properties of contrasts are contained in the following three theorems.

Theorem 15-2

An unbiased estimate of a contrast is given by $\hat{\psi} = \alpha_1 \bar{X}_1 + a_2\bar{X}_2 + \cdots + a_K\bar{X}_K$.

Proof. By definition, $\hat{\psi} = a_1\bar{X}_1 + a_2\bar{X}_2 + \cdots + a_K\bar{X}_K$. By Theorems 6-1 and 7-3,

$$\begin{aligned} E(\hat{\psi}) &= E(a_1\bar{X}_1) + E(a_2\bar{X}_2) + \cdots + E(a_K\bar{X}_K) \\ &= a_1 E(\bar{X}_1) + a_2 E(\bar{X}_2) + \cdots + a_K E(\bar{X}_K) \\ &= a_1\mu_1 + a_2\mu_2 + \cdots + a_K\mu_K \\ &= \psi \end{aligned}$$

This completes the proof.

Thus, the simple contrast $\hat{\psi} = \bar{X}_1 - \bar{X}_2$ used to compare two population means is an unbiased estimate of their difference, $\psi = \mu_1 - \mu_2$. For the study on Parkinson's law mentioned in the quotation at the beginning of this chapter, the contrast for the 5-minute versus the 15-minute treatment groups is given by $\hat{\psi} = \bar{X}_5 - \bar{X}_{15} = 5 - 8 = -3$. This suggests that the group given 15 minutes for preparation required 3 extra minutes to accomplish the same task that the others performed in 5 minutes.

Theorem 15-3

If samples are selected from independent universes, the variance of a contrast is given by

$$\operatorname{Var}(\hat{\psi}) = \sigma^2\left[\frac{a_1^2}{N_1} + \frac{a_2^2}{N_2} + \cdots + \frac{a_K^2}{N_K}\right]$$

Proof. By definition, $\hat{\psi} = a_1 \bar{X}_1 + a_2 \bar{X}_2 + \cdots + a_K \bar{X}_K$. By Theorems 6-4 and 7-4,

$$\begin{aligned}
\operatorname{Var}(\hat{\psi}) &= \operatorname{Var}(a_1 \bar{X}_1) + \operatorname{Var}(a_2 \bar{X}_2) + \cdots + \operatorname{Var}(a_K \bar{X}_K) \\
&= a_1^2 \operatorname{Var}(\bar{X}_1) + a_2^2 \operatorname{Var}(\bar{X}_2) + \cdots + a_K^2 \operatorname{Var}(\bar{X}_K) \\
&= a_1^2 \frac{\sigma^2}{N_1} + a_2^2 \frac{\sigma^2}{N_2} + \cdots + a_K^2 \frac{\sigma^2}{N_K} \\
&= \sigma^2 \left[\frac{a_1^2}{N_1} + \frac{a_2^2}{N_2} + \cdots + \frac{a_K^2}{N_K} \right]
\end{aligned}$$

This completes the proof.

Since σ^2 is generally unknown, MSW is substituted for σ^2 so that an unbiased estimate of the variance of a contrast is given by

$$\mathrm{SE}_{\hat{\psi}}^2 = \mathrm{MSW} \left[\frac{a_1^2}{N_1} + \frac{a_2^2}{N_2} + \cdots + \frac{a_K^2}{N_K} \right]$$

For the contrast $\hat{\psi} = \bar{X}_1 - \bar{X}_2$, the squared standard error is given by the familiar

$$\mathrm{SE}_{\hat{\psi}}^2 = S_p^2 \left[\frac{1}{N_1} + \frac{1}{N_2} \right]$$

With these definitions and notations it is now possible to state Scheffé's theorem. The proof of this theorem is beyond the scope of this book.

Theorem 15-4

Consider the set of all possible contrasts of the form $\psi = a_1 \mu_1 + \cdots + a_K \mu_K$. The probability that confidence intervals of the form

$$\hat{\psi} - S\mathrm{SE}_{\hat{\psi}} < \psi < \hat{\psi} \pm S\mathrm{SE}_{\hat{\psi}}$$

are simultaneously true statements is $(1 - \alpha)$ if

$$S = \sqrt{(K-1)\, F_{K-1, N-K}(1-\alpha)}.$$

An extremely important consequence of this theorem is that if H_0 is rejected by the F test with a probability of a type I error set equal to α, then there is at least one confidence interval for the complete set of possible ψ's that does not contain 0, provided that the multiplier for the $\mathrm{SE}_{\hat{\psi}}$ is set equal to S. This means that if confidence intervals such as

$$\psi = \hat{\psi} \pm S\mathrm{SE}_{\hat{\psi}}$$

are computed, then eventually one will be found that does not include 0. Thus, according to Scheffé's theorem, it is possible to study observed data and search out meaningful contrasts and confidence intervals to help identify possible reasons

for the rejection of a tested hypothesis. In essence, this justifies "data snooping" after the fact. In this sense, these contrasts are sometimes called *post hoc* comparisons.

For the data of Table 14-1,

$$S = \sqrt{(K-1)F_{K-1,N-K}(1-\alpha)} = \sqrt{(4-1)F_{3,20}(.95)}$$
$$= \sqrt{3(3.10)} = 3.05$$

The number of different two-parameter contrasts for the data of Table 14-1 is given by $\binom{4}{2} = 6$. The computations associated with each of the six pairwise contrasts are summarized in Table 15-2. According to these results, treatment 1 differs from

Table 15-2. Computations for the pair-wise simple comparisons of the data of Table 14-1.

Comparison	*Value of* $\hat{\psi}$	$\mathrm{SE}^2_{\hat{\psi}}$	*Lower limit*	*Upper limit*	*Decision*
$\psi_1 = \mu_1 - \mu_2$	12.17 − 17.43 = −5.26	$46.37(\frac{1}{6} + \frac{1}{7}) = 14.35$	−16 82	+ 6 30	Not significant
$\psi_2 = \mu_1 - \mu_3$	12.17 − 28.00 = −15.83	$46.37(\frac{1}{6} + \frac{1}{6}) = 15.46$	−27 82	− 3.84	Significant
$\psi_3 = \mu_1 - \mu_4$	12.17 − 28.80 = −16.63	$46.37(\frac{1}{6} + \frac{1}{5}) = 17.00$	−29.20	− 4.06	Significant
$\psi_4 = \mu_2 - \mu_3$	17.43 − 28.00 = −10.57	$46.37(\frac{1}{7} + \frac{1}{6}) = 14\ 35$	−22.13	+ 99	Not significant
$\psi_5 = \mu_2 - \mu_4$	17.43 − 28.80 = −11.37	$46.37(\frac{1}{7} + \frac{1}{5}) = 15.90$	−23 54	+ .80	Not significant
$\psi_6 = \mu_3 - \mu_4$	28.00 − 28.80 = −.80	$46\ 37(\frac{1}{6} + \frac{1}{5}) = 17.00$	−13.37	+11 77	Not significant

treatments 3 and 4. Note that inspection of the data suggests that the effects for treatments 3 and 4 are identical and that perhaps together they may be different from treatment 2. This *post hoc* hypothesis can be easily tested by means of the following contrast:

$$\psi_7 = \mu_2 - (\tfrac{6}{11}\mu_3 + \tfrac{5}{11}\mu_4)$$

An unbiased estimate of ψ_7 is given by

$$\hat{\psi}_7 = \bar{X}_2 - (\tfrac{6}{11}\bar{X}_3 + \tfrac{5}{11}\bar{X}_4)$$
$$= 17.43 - (\tfrac{6}{11}(28.00) + \tfrac{5}{11}(28.00))$$
$$= -10.93$$

and an unbiased estimate of its variance is given by

$$\mathrm{SE}^2_{\hat{\psi}_7} = \mathrm{MSW}\left[\frac{1}{N_2} + \frac{(-\frac{6}{11})^2}{N_3} + \frac{(-\frac{5}{11})^2}{N_4}\right]$$
$$= 46.37\left[\frac{1}{7} + \frac{\frac{36}{121}}{6} + \frac{\frac{25}{121}}{5}\right]$$
$$= 10.84$$

The confidence interval for ψ_7 is given by

$$-10.93 - 3.05\sqrt{10.84} < \psi_7 < -10.93 + 3.05\sqrt{10.84}$$
$$-20.96 < \psi_7 < -.90$$

Since 0 is not included in this interval, there is reason to believe that treatment 2 differs collectively from 3 and 4.

As a summary statement, one can say that treatments 1 and 2 are similar and treatments 3 and 4 are similar. This suggests that large penalties for wrong choices limit or reduce guessing.

Note that seven confidence intervals have been investigated in this study. The probability of at least one type I error in the set of seven intervals is guaranteed by Scheffé's theorem to be equal to .05. If, instead of S, as defined by Scheffé's theorem, one had used a t value at .05 for each of the seven confidence intervals, then the probability of at least one type I error would be especially high. In fact, one can show that the overall probability of a type I error would be at most equal to $\alpha_0 = 7(.05) = .35$.

In particular, if Q intervals are investigated, the probability of at least one type I error is at most equal to $\alpha_0 \leqslant Q\alpha$, where α is the probability of a type I error associated with each interval. This illustrates the great beauty and power of Scheffé's theorem. It is possible to investigate as many intervals as desired and still control the probability of a type I error at any specified α level. One other advantage associated with this theorem is that unlimited data snooping via contrasts is possible. The importance of this statement will be well appreciated by all researchers.

15-6 TUKEY'S METHOD OF MULTIPLE PAIRWISE CONTRASTS

One method that has considerable efficiency for *pairwise* contrasts with *equal* sample sizes is Tukey's method of multiple comparisons. For this method, one replaces the Scheffé S coefficient with a Tukey T coefficient, which is defined as $T = (1/\sqrt{2})q$. To determine T, one enters Table A-11 with $\nu_1 + 1 = K$ and $\nu_2 = N - K$ and reads the corresponding value of q. From this, T is determined. Thus, for the data of Table 14-2, $\nu_1 + 1 = K = 6$ and $\nu_2 = N - K = 48 - 8 = 40$. One reads in Table A-11 that $q = 4.23$ for $\alpha = .05$ and that $T = (1/\sqrt{2})(4.23) = 2\ 99$. The corresponding Scheffé coefficient is given by

$$S = \sqrt{(K-1)F_{K-1,N-K}(1-\alpha)} = \sqrt{5(2.45)} = \sqrt{12.25} = 3.50$$

This shows that in this case the Scheffé interval will be approximately one standard deviation wider than the corresponding Tukey interval.

Since the F test is the K-sample extension of the t test, one might think that the F test can be shown to be identical to the t test when $K = 2$. This connection with the t test was indicated in Section 14.8. Because of this, one might expect the Scheffé

intervals to reduce to the simple confidence interval for the difference between two means when $K = 2$. When $K = 2$,

$$\begin{aligned} S &= \sqrt{(K-1)\,F_{K-1,N-K}(1-\alpha)} \\ &= \sqrt{(2-1)\,F_{1,N-2}(1-\alpha)} = \sqrt{F_{1,N-2}(1-\alpha)} \\ &= t_{N-2}\left(\frac{\alpha}{2}\right) \end{aligned}$$

Since $N - 2 = N_1 + N_2 - 2$ and $\mathrm{MSW} = S_p^2$, the confidence interval for $\psi = \mu_1 - \mu_2$ is given by

$$(\bar{X}_1 - \bar{X}_2) \pm t_{N_1+N_2-2}\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_p^2}{N_1} + \frac{S_p^2}{N_2}}$$

as noted in Section 11-8.

15-7 THE F^* TEST THE K-SAMPLE ANALOG TO THE F TEST WHEN POPULATION VARIANCES ARE UNEQUAL

When the variances are unequal and the parent populations are normal, one can perform a test that is the K-sample analog to the Welch–Aspin two-sample t test. When $K = 2$, this test reduces to the two-sample Welch–Aspin test. As in the Welch–Aspin test, one substitutes into the classical test statistic the individual sample variances in place of the pooled estimate of the variance and then computes an effective number of degrees of freedom to determine the approximate distribution of the resulting test statistic.

If S_p^2 is substituted for MSW, the classical test statistic is given by

$$F = \frac{\mathrm{MSB}}{\mathrm{MSW}} = \frac{1}{K-1}\sum_{k=1}^{K}\frac{N_k(\bar{X}_k - \bar{X})^2}{S_p^2} = \frac{1}{K-1}\left[\sum_{k=1}^{K}\frac{N_k}{S_p^2}(\bar{X}_k - \bar{X})^2\right]$$

The Welch–Aspin analog to this statistic is given by

$$F^* = \frac{1}{K-1}\left[\sum_{k=1}^{K}\frac{N_k}{S_k^2}(\bar{X}_k - \bar{X}^*)^2\right] = \frac{1}{K-1}\sum_{k=1}^{K}\hat{W}_k(\bar{X}_k - \bar{X}^*)^2$$

where

$$\hat{W}_k = \frac{N_k}{S_k^2} \qquad \text{and} \qquad \bar{X}^* = \frac{\sum_{k=1}^{K}\hat{W}_k\bar{X}_k}{\sum_{k=1}^{K}\hat{W}_k}$$

If the hypothesis H_0 $\mu_1 = \mu_2 = \cdot \cdot = \mu_K = \mu_0$ is true, then F^* has an approximate F distribution with $\nu_1 = K - 1$ and $\nu_2 = \nu_2^*$, where

$$\nu_2^* = \frac{K^2 - 1}{3\Lambda}$$

$$\Lambda = \sum_{k=1}^{K} \frac{1}{N_k - 1}\left(1 - \frac{\hat{W}_k}{W}\right)^2$$

$$W = \sum_{k=1}^{K} \hat{W}_k$$

The derivation of this test statistic is beyond the methods presented in this book. However, it is easy to show that $\bar{X}^*$ and F^* reduce to the usual $\bar{X}$ and F statistic when MSW is substituted for each S_k^2.

Theorem 15-5

The estimate of the mean $\bar{X}^*$ reduces to the usual estimate of the mean $\bar{X}$ when the pooled estimate of the variance is substituted for each individual variance.

Proof. By definition,

$$\bar{X}^* = \frac{\sum_{k=1}^{K} \hat{W}_k \bar{X}_k}{\sum_{k=1}^{K} \hat{W}_k}$$

$$= \frac{\sum_{k=1}^{K} (N_k/S_k^2)\, \bar{X}_k}{\sum_{k=1}^{K} N_k/S_k^2}$$

Replacing each S_k^2 by S_p^2, we have

$$\bar{X}^* = \frac{\sum_{k=1}^{K} (N_k/S_p^2)\, \bar{X}_k}{\sum_{k=1}^{K} N_k/S_p^2}$$

$$= \frac{1/S_p^2 \sum_{k=1}^{K} N_k \bar{X}_k}{1/S_p^2 \sum_{k=1}^{K} N_k}$$

$$= \frac{\sum_{k=1}^{K} N_k \bar{X}_k}{\sum_{k=1}^{K} N_k}$$

$$= \frac{1}{N} \sum_{k=1}^{K} N_k \bar{X}_k$$

This completes the proof.

When the sample variances are equal, $\bar{X}^*$ is simply the weighted average of the individual treatment means. It is worth noting that while $\bar{X}^*$ is generally a biased estimator of μ, its sampling distribution is more compact than that of $\bar{X}$. In this sense it is somewhat more efficient as an estimator of μ than is $\bar{X}$.

Theorem 15-6

F^* reduces to F, when the assumptions for the F test are satisfied.

Proof. If $\bar{X}$ is substituted for $\bar{X}^*$ and if S_p^2 is substituted for each S_k^2, then

$$F^* = \frac{1}{K-1}\sum_{k=1}^{K}\frac{N_k}{S_k^2}(\bar{X}_k - \bar{X}^*)^2$$

$$= \frac{1}{K-1}\sum_{k=1}^{K}\frac{N_k}{S_p^2}(\bar{X}_k - \bar{X})^2$$

$$= \frac{1}{S_p^2}\left[\frac{1}{K-1}\sum_{k=1}^{K} N_k(\bar{X}_k - \bar{X})^2\right] = \frac{\text{MSB}}{\text{MSW}}$$

This completes the proof.

As an illustration of the use of this test, reconsider the data of Table 14-1. The computations for the F^* test are summarized in Table 15-3. For the data of Table 15-3,

Table 15-3. Computation table for the determination of F^*.

Treatment	N_k	S_k^2	$\hat{W}_k$	$\bar{X}_k$	$\hat{W}_k\bar{X}_k$	$(\bar{X}_k - \bar{X}^*)$
1	6	13 37	4488	12.17	5 4619	−4 56
2	7	43 29	.1617	17.43	2 8184	+.70
3	6	71.60	0838	28 00	2.3464	+11.27
4	5	60 70	0824	28.80	2.3731	+12.07
			7767		12.9998	

$$\bar{X}^* = \frac{\sum_{k=1}^{K}\hat{W}_k\bar{X}_k}{\sum_{k=1}^{K}\hat{W}_k} = \frac{12.9998}{.7767} = 16.73$$

$$F^* = \frac{1}{4-1}[.4488(-4.56)^2 + .1617(.70)^2 + .0838(11.27)^2 + .0824(12.07)^2]$$

$$= 10.68$$

Note that $F = 8.32$ has increased to $F^* = 10.68$. This increase in test value is characteristic of this test but is not necessarily indicative of greater power.

The computations required for the determination of Λ are summarized in Table 15-4. For these data, $\Lambda = .4992$ and $\nu_2^* = (4^2 - 1)/3(.4992) = 10.016 \simeq 10$ Note

Table 15-4. Computations required for the determination of Λ.

Treatment	N_k	$\frac{1}{N_k - 1}$	$\hat{W}_k$	$\frac{\hat{W}_k}{W}$	$1 - \frac{\hat{W}_k}{W}$	$\left(1 - \frac{\hat{W}_k}{W}\right)^2$	$\frac{1}{N_k - 1}\left(1 - \frac{\hat{W}_k}{W}\right)^2$
1	6	$\frac{1}{5}$ = 2000	4488	.5778	.4222	.1783	0357
2	7	$\frac{1}{6}$ = .1667	.1617	.2081	.7919	6271	.1045
3	6	$\frac{1}{5}$ = .2000	.0838	.1079	.8921	7958	1592
4	5	$\frac{1}{4}$ = 2500	0824	.1061	.8939	7991	1998
							4992

that the number of degrees of freedom has been reduced from an actual 20 to an effective 10. This reduction in power has occurred because of the variability that exists between the sample variances. If the sample variances had been quite discrepant, then the reduction in degrees of freedom would have been much larger. With $\nu_1 = 3$ and $\nu_2^* = 10$, the 95 percent decision rule for the rejection of the hypothesis of equal expected values is to reject H_0 if $F^* > F_{3,10}(.95) = 3.71$. In this case, H_0 is rejected since $10.68 > 3.71$. Therefore, there is reason to believe that in the population of scores, the expected values for the four treatment conditions are indeed different.

15-8 MULTIPLE CONTRASTS FOR THE F^* TEST

The results of Scheffé's theorem can be extended to *post hoc* comparison methods following the F^* test simply by replacing S with

$$S^* = \sqrt{(K-1)\,F_{K-1,\nu_2^*}(1-\alpha)}$$

where ν_2^* is as defined in Section 15-7. Furthermore, in place of the pooled estimate of the common variance, the individual sample variances should be substituted.

15-9 EXPLAINED VARIANCE IN THE ANALYSIS-OF-VARIANCE MODEL

In Section 13-6, it was seen that when $H_0\colon \mu_1 = \mu_2$ was rejected on the basis of a t test, it followed that conditional probability statements about the criterion variable would lead to unequal probabilities. The same is true when $H_0\colon \mu_1 = \mu_2 = \cdots$

$= \mu_K = \mu_0$ is rejected on the basis of the F test. While one could compute conditional probabilities of interest, this practice is generally bypassed and in its place a measure of the strength of the association or relationship is computed. This measure, denoted by $\hat{\omega}^2$, was introduced in Section 14-5 and is further examined here.

In the analysis of variance model, it was seen that

$$\text{SST} = \text{SSB} + \text{SSW}$$

Upon division by SST, this result reduces to

$$1 = \frac{\text{SSB}}{\text{SST}} + \frac{\text{SSW}}{\text{SST}}$$

When H_0 is true, SSB is quite small so that SSW is almost identical to SST. Since SSW is a measure of the variability in the criterion variable that is unrelated to the treatment or experimental conditions of the study, it is said to be the sum of squares that is unexplained. In this sense, SSW/SST $\times$ 100 is a percentage measure of unexplained variability, while SSB/SST $\times$ 100 is a percentage measure of explained variance. This measure will equal 0 when H_0 is true or when all conditional distributions of the study are identical. Thus, instead of computing $P(X > X_0 | \mu_k)$ for all k and various values of X_0, one can compute a summary measure of the strength of the association. This measure is given by

$$\hat{\omega}^2 = \frac{\text{sum of squares explained}}{\text{sum of squares total}}$$

which for an analysis-of-variance design reduces to

$$\hat{\omega}^2 = \frac{(K-1)\,\text{MSB}}{(N-1)\,\text{MST}}$$

For the risk-taking problem, it was seen that

$$\hat{\omega}^2 = .55$$

Since the range of $\hat{\omega}^2$ extends from 0 to 1, it follows that the manipulated variable of the study, the directions given to the students concerning the risk of wrong answers, is a strong determinant of the subjects' behavior.

15-10 TESTS AND CONFIDENCE INTERVALS FOR THE QUANTITATIVE K-SAMPLE MODEL

The statistical tests and confidence-interval procedures introduced in this chapter for the K sample quantitative analysis-of-variance model are summarized in Table 15-5.

Table 15-5. *K*-sample tests and confidence intervals for quantitative model on population centers.

Case	*Hypotheses*	*Test statistic*	*Confidence intervals*	*Assumptions*
13	$H_0: \mu_1 = \mu_2 = \cdots = \mu_K$ $H_1: H_0$ is false	$F = \dfrac{\text{MSB}}{\text{MSW}} = \dfrac{1/(K-1)\sum_{k=1}^{K} N_k(\bar{X}_k - \bar{X})^2}{1/(N-1)\sum_{k=1}^{K}(N_k - 1)S_k^2}$ $\nu_1 = K - 1$ $\nu_2 = N - K$	$\mu_{k1} - \mu_{k2} = (\bar{X}_{k_1} - \bar{X}_{k_2}) \pm S\sqrt{\dfrac{\text{MSW}}{N_{k_1}} + \dfrac{\text{MSW}}{N_{k_2}}}$ $S = \sqrt{(K-1)F_{K-1,N-K}(1-\alpha)}$ $k_1, k_2 = 1, 2, 3, 4, \ldots, K$ $k_1 < k_2$	1. Independence between samples 2 Independence within samples 3 Normality 4. Variances unknown but equal
14	$H_0: \mu_1 = \mu_2 = \cdot\ \cdot = \mu_K$ $H_1: H_0$ is false	$F^* = \dfrac{1}{K-1}\sum_{k=1}^{K} \hat{W}_k(\bar{X}_k - \bar{X}^*)^2$ where $\hat{W}_k = \dfrac{N_k}{S_k^2} \qquad W = \sum_{k=1}^{K} \hat{W}_k$ $\bar{X}^* = \dfrac{\sum_{k=1}^{K} \hat{W}_k \bar{X}_k}{\sum_{k=1}^{K} \hat{W}_k}$ $\nu_1 = K - 1$ $\nu_2^* = \dfrac{K^2 - 1}{3\Lambda}$ $\Lambda = \sum_{k=1}^{K} \dfrac{1}{N_k - 1}\left(1 - \dfrac{\hat{W}_k}{W}\right)^2$	$\mu_{k1} - \mu_{k2} = (\bar{X}_{k_1} - \bar{X}_{k_2}) \pm S^*\sqrt{\dfrac{S_{k1}^2}{N_{k_1}} + \dfrac{S_{k2}^2}{N_{k_2}}}$ $S^* = \sqrt{(K-1)F_{K-1,\nu_2^*}(1-\alpha)}$ $k_1, k_2 = 1, 2, \ldots, 6$ $k_1 < k_2$	1 Independence between samples 2 Independence within samples 3 Normality 4. Variances unknown and unequal

15-11 SUMMARY

In this chapter, the frequent test of hypothesis H_0 $\mu_1 = \mu_2 = \cdots = \mu_K$ was discussed and presented. This test, a simple extension of the t test, is based upon the test statistic $F = \text{MSB}/\text{MSW}$, where

$$\text{MSB} = \frac{1}{K-1} \sum_{k=1}^{K} N_k(\bar{X}_k - \bar{X})^2$$

and

$$\text{MSW} = \frac{1}{N-K} \sum_{k=1}^{K} (N_k - 1)\, S_k^2$$

When the underlying variable has a normal or near-normal distribution with equal variance in each sampled universe, it follows that F has a sampling distribution that is F, with parameters $\nu_1 = K - 1$ and $\nu_2 = N - K$, provided that independent random samples are selected from each universe.

While the derivation of the F statistic was lengthy and tortuous, it is possible to derive the test with ease, using the theory of Chapter 10. As was shown in Theorem 10-8, we have for the kth sample

$$\sum_{i=1}^{N_k} \left(\frac{X_{ik} - \mu}{\sigma}\right)^2 = \frac{(N_k - 1)\, S_k^2}{\sigma^2} + \frac{N_k(\bar{X}_k - \mu)^2}{\sigma^2}$$

with

$$\chi^2_{N_k} = \chi^2_{N_k - 1} + \chi^2_1$$

By repeated applications of Theorem 10-8 over K samples, we have

$$\sum_{k=1}^{K} \sum_{i=1}^{N_k} \left(\frac{X_{ik} - \mu}{\sigma}\right)^2 = \sum_{k=1}^{K} \frac{(N_k - 1)\, S_k^2}{\sigma^2} + \sum_{k=1}^{K} \frac{N_k(\bar{X}_k - \mu)^2}{\sigma^2}$$

$$= \frac{\text{SSW}}{\sigma^2} + \sum_{k=1}^{K} \frac{N_k(\bar{X}_k - \mu)^2}{\sigma^2}$$

with

$$[\chi^2_{N_1} + \chi^2_{N_2} + \cdots + \chi^2_{N_K}] = [\chi^2_{N_1 - 1} + \chi^2_{N_2 - 1} + \cdots + \chi^2_{N_K - 1}] + [1 + 1 + \cdots + 1]$$

or

$$\chi^2_N = \chi^2_{N-K} + \chi^2_K$$

Employing the binomial expansion to $(\bar{X}_k - \mu)$ by adding and subtracting μ_k, we can show that

$$\sum_{k=1}^{K} \frac{N_k(\bar{X}_k - \mu)^2}{\sigma^2} = \sum_{k=1}^{K} \frac{N_k(\bar{X}_k - \bar{x})^2}{\sigma^2} + \frac{N(\bar{X} - \mu)^2}{\sigma^2}$$

$$= \frac{\text{SSB}}{\sigma^2} + \frac{N(\bar{X} - \mu)^2}{\sigma^2}$$

with

$$\chi_K^2 = \chi_{K-1}^2 + \chi_1^2$$

Putting these results together, we have

$$\sum_{k=1}^{K} \sum_{i=1}^{N_k} \left(\frac{X_{ik} - \mu}{\sigma}\right)^2 = \frac{\text{SSW}}{\sigma^2} + \frac{\text{SSB}}{\sigma^2} + \frac{N(\bar{X} - \mu)^2}{\sigma^2}$$

with

$$\chi_N^2 = \chi_{N-K}^2 + \chi_{K-1}^2 + \chi_1^2$$

According to the definition of F given in Theorem 14-1,

$$F = \frac{\dfrac{\chi_{K-1}^2}{K-1}}{\dfrac{\chi_{N-K}^2}{N-K}} = \frac{\dfrac{\text{SSB}/\sigma^2}{K-1}}{\dfrac{\text{SSW}/\sigma^2}{N-K}}$$

$$= \frac{\text{SSB}/(K-1)}{\text{SSW}/(N-K)}$$

$$= \frac{\text{MSB}}{\text{MSW}}$$

Fortunately, the assumptions of normality and common variance can be somewhat relaxed when studying behavioral variables. If the sample sizes are large, there is good evidence that F will approach the theoretical F distribution even if the variable is not normal. If the variances are unequal, and if equal sample sizes are selected from each universe, then the approximation to the theoretical F distribution is also quite good. This suggests that equal-sized samples should be used for behavioral research whenever possible.

Whereas rejection of H_0 in the two-sample problem indicates that $\mu_1 \neq \mu_2$, rejection of H_0 in the K-sample problem is not very informative since one does not know which expected values are different from one another. To locate the sources of rejection of the tested hypothesis, a researcher must rely on some *post hoc* procedure that generates confidence intervals that may be examined for the inclusion

or exclusion of 0. The most popular procedure, developed by Scheffé and based on linear contrasts ψ, is dependent upon computing confidence intervals of the form

$$\hat{\psi} - S(\mathrm{SE})_{\hat{\psi}} < \psi < \hat{\psi} + S(\mathrm{SE}_{\hat{\psi}})$$

where

1. $\psi = a_1\mu_1 + a_2\mu_2 + \cdots + a_K\mu_K$
2. $a_1 + a_2 + \cdots + a_K = 0$
3. $\hat{\psi} = a_1\bar{X}_1 + a_2\bar{X}_2 + \cdots + a_K\bar{X}_K$
4. $S^2 = (K-1)F_{K-1,N-K}(1-\alpha)$
5. $\mathrm{SE}^2_{\hat{\psi}} = \mathrm{MSW}\left[\frac{a_1^2}{N_1} + \frac{a_2^2}{N_2} + \cdots + \frac{a_K^2}{N_K}\right]$

Scheffé has shown that the probability of making at least one type I error in the complete set of confidence intervals of interest is equal to α, provided that S is used as the multiplying coefficient applied to the standard errors of the contrasts of interest.

If a researcher believes that the population variances are unequal and if equal-sized samples cannot be obtained, then consideration should be given to the Welch–Aspin analog to the F test by using

$$F^* = \frac{1}{K-1}\sum_{k=1}^{K}\hat{W}_k(\bar{X}_k - \bar{X}^*)^2$$

where

$$\bar{X}^* = \frac{\sum_{k=1}^{K}\hat{W}_k\bar{X}_k}{\sum_{k=1}^{K}\hat{W}_k} \qquad \text{and} \qquad \hat{W}_k = \frac{1}{\mathrm{SE}^2_{\bar{X}_k}}$$

which, when H_0 is true and when the $\bar{X}_k$ are based upon independent random samples, has a sampling distribution that can be approximated by an F variable with

$$\nu_1 = K-1 \qquad \text{and} \qquad \nu_2^* = \frac{K^2-1}{3\Lambda}$$

where

$$\Lambda = \sum_{k=1}^{K}\frac{1}{N_k-1}\left(1-\frac{\hat{W}_k}{W}\right)^2 \qquad \text{and} \qquad W = \sum_{k=1}^{K}\hat{W}_k$$

When H_0 is rejected, sources of the rejection can be investigated by using

$$S^{*2} = (K-1)F_{K-1,\nu_2^*}(1-\alpha)$$

in place of S in the corresponding Scheffé intervals.

Generally, the defining formulas for MSB and MSW are bypassed in the actual performance of an analysis of variance since computing formulas can reduce the amount of arithmetic by a considerable degree. If $T_{.k}$ represents the total of the observed elements in the kth sample, then

1. $$\text{MSB} = \frac{\text{SSB}}{K-1} = \frac{1}{K-1}\left[\sum_{k=1}^{K} \frac{T_{.k}^2}{N_k} - \frac{T_{..}^2}{N}\right] = \frac{1}{K-1}[\text{III}_C - \text{II}]$$

2. $$\text{MSW} = \frac{\text{SSW}}{N-K} = \frac{1}{N-K}\left[\sum_{k=1}^{K}\sum_{i=1}^{N_k} X_{ik}^2 - \sum_{k=1}^{K} \frac{T_{.k}^2}{N_k}\right] = \frac{1}{N-K}[\text{I} - \text{III}_C]$$

3. $$\text{MST} = \frac{\text{SST}}{N-1} = \frac{1}{N-1}\left[\sum_{k=1}^{K}\sum_{i=1}^{N_k} X_{ik}^2 - \frac{T^2}{N}\right] = \frac{1}{N-1}[\text{I} - \text{II}]$$

In terms of K, N, I, II, and III_C, the statistics for an analysis-of-variance design can be represented in an analysis-of-variance table, as shown in Table 15-6.

Table 15-6. The general analysis-of-variance table.

Source of variance	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F ratio*	*Expected mean square*
Between groups	$K-1$	$\text{III}_C - \text{II}$	MSB	$\frac{\text{MSB}}{\text{MSW}}$	$\sigma^2 + \frac{1}{K-1}\sum_{k=1}^{K} N_k(\mu_k - \mu)^2$
Within groups	$N-K$	$\text{I} - \text{III}_C$	MSW		σ^2
Total	$N-1$	$\text{I} - \text{II}$			

Finally, it should be noted that the analysis of variance is one of the most frequently used statistical tools of behavioral research. It can be extended in a number of ways and be used to test hypotheses relevant to behavioral scientists. Because of this, it is recommended that students who are contemplating a career in behavioral research take time to obtain more training in the use of this important statistical model.

EXERCISES

***15-1.** In a study in which success in learning elementary arithmetic was the primary object, a group of third-grade children were asked "How well do you like school?" Answer

choices were. {very well, fairly well, not very well, not at all}. Later, the same children were given an arithmetic reasoning test. The scores according to response to the question were as follows:

Response to question			
VERY WELL	FAIRLY WELL	NOT VERY WELL	NOT AT ALL
38	40	18	16
39	60	33	27
26	37	19	19
40	19	25	30
45	37	37	
53	26	28	
	31		
	42		
	45		

Is performance related to attitude?

***15-2.** What have you assumed in answering the question of Exercise 15-1? Are these assumptions satisfied?

***15-3.** Assign a Likert scale to individual attitudes toward the integration of junior high schools and determine whether the data of Exercise 3-10 are indicative of uniform attitudes toward integration of elementary schools. What have you assumed? Are these assumptions reasonable? Explain.

15-4. In a survey of families living in San Diego, Los Angeles, Oakland, and San Francisco, the following hypothetical statistics were obtained concerning the amount of money spent per week on food for families with four members:

City	*Sample size*	*Average*	*Standard deviation*
San Diego	80	42 75	7.23
Los Angeles	200	37 28	6.00
Oakland	140	35.69	4.65
San Francisco	60	39.41	5.92

What do these statistics suggest about the cost of food in these cities?

15-5. There are some researchers who believe that the performance of the F test is a waste of time and that one should go directly to the Scheffé method of multiple contrasts. Comment on this position. Is it a valid point of view? Explain.

***15-6.** One can reduce the arithmetic in the analysis by using the computational formulas. Further reduction can be achieved by subtracting a constant from all of the data, since by Theorem 8-2, it would follow that $S_X^2 = S_Y^2$, where $Y = X - a$. Do the F test on the data of Table 14-1, after subtracting 21 from each of the observed values

***15-7.** Analyze the data of Exercise 8-5 as an F test and as a t test and compare the results. What is the relationship between the computed t and F values?

***15-8.** Analyze the data of Exercise 8-7 in light of the methods presented in this chapter. Explain what you observe.

***15-9.** Analyze the data of Exercise 11-2 in light of the methods presented in this chapter

***15-10.** Analyze the data of Table 14-2 in light of the methods presented in this chapter

16 HYPOTHESIS TESTING IN THE *K*-SAMPLE QUALITATIVE CHI-SQUARE MODEL

MUSIC HATH CHARMS . . .

Twelve hours a day for nearly two months, three groups of albino rats at a Texas Tech University laboratory were given some musical entertainment. One group of newborn rat pups was exposed to selections from Mozart—*The Magic Flute, Symphonies 40* and *41*, the *Violin Concerto No 5*. A second group audited an equivalent daily dose of Arnold Schoenberg—*Pierrot Lunaire, Verklarte Nacht,* and *Kol Nidre*, among other compositions. The third set of rats, appointed as a control, heard nothing but the whirring of a ventilation fan.

At the end of this calculated bombardment, the three colonies were granted a 15-day respite from all music. Then they entered cages which allowed them, by tripping electric circuits, to opt either for Mozart or Schoenberg—in both cases, compositions they had not heard before—or to listen to nothing but the fan. The results should be encouraging to Mozart buffs. The rats exposed to his music during their compulsory concerts overwhelmingly tuned in on him The group indoctrinated by Schoenberg split almost evenly between him and Mozart—as did the control group, which was unfamiliar with both composers.

16-1 THE TWO-SAMPLE BINOMIAL TEST AS A CHI-SQUARE TEST

In Sections 11-11 and 13-5, the test of hypothesis and the confidence-interval procedure for testing the differences in two binomial parameters were presented. As was shown, the test statistic for the test of H_0. $\theta = p_1 - p_2$ versus the alternative $H_1 \cdot H_0$ is false is given by

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0\hat{q}_0/N_1 + \hat{p}_0\hat{q}_0/N_2}}$$

where

$$\hat{p}_0 = \frac{N_1\hat{p}_1 + N_2\hat{p}_2}{N_1 + N_2}$$

and the $(1 - \alpha)$ percent confidence interval for $\theta = p_1 - p_2$ is given by

$$(\hat{p}_1 - \hat{p}_2) - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1} + \frac{\hat{p}_2\hat{q}_2}{N_2}} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1} + \frac{\hat{p}_2\hat{q}_2}{N_2}}$$

provided that p_1 and p_2 are estimated from independent random samples and the sample sizes are large enough so that the normal approximation is adequate.

Since the model that gave rise to these specific results cannot be extended with ease, it will be useful to consider an alternative procedure that gives rise to exactly the same results but in a different computational form. The reasons for considering this alternative method is that it can be easily extended to hypothesis-testing procedures and confidence-interval statements that are valid for more than two populations. For this discussion, assume that the data has been tabulated and reported as shown in Table 16-1.

Table 16-1. Table of frequencies for the outcomes on two independent samples from binomial populations.

Outcome	*Sample 1*	*Sample 2*	*Total frequencies*
A	X_{11}	X_{21}	X_1
$\bar{A}$	X_{12}	X_{22}	X_2
Total frequencies	N_1	N_2	N

In terms of the notation of Table 16-1, $\hat{p}_1 = X_{11}/N_1$, $\hat{p}_2 = X_{21}/N_2$, and $\hat{p}_0 = X_1/N$ are unbiased estimates of p_1, p_2, and p_0. If H_0 is true, these three estimates of p_0 will vary only within the expected range of chance variation.

While Z can be used to test the hypothesis of equal p values, many researchers use another statistic proposed by Karl Pearson at the turn of the century This "classical test" statistic is algebraically equivalent to Z as a test statistic. For Pearson's statistic, one first determines a new table of estimated frequencies from the table of observed frequencies and then computes

$$X^2 = \sum_{k=1}^{2} \sum_{i=1}^{2} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$$

where X_{ki} is the observed frequency in the ith subset of population k and $\hat{E}(X_{ki})$ is the estimated expected frequency in the ith subset of population k. This statistic is then related to the chi-square distribution with $\nu = 1$.

For Table 16-1, the estimated frequencies can be portrayed as shown in Table 16-2. For this table, $\hat{p}_0 = X_1/N$ and $\hat{q}_0 = X_2/N$. It should be noted that Tables 16-1 and 16-2 have the same marginal frequencies. If the test of H_0 is performed as a Karl Pearson chi-square test and if it is noted that the marginal frequencies are not

Table 16-2. Estimated expected frequencies for the classical chi-square test of Karl Pearson. Estimated frequencies determined from the observed frequencies of Table 16-1.

Outcome	*Sample 1*	*Sample 2*	*Total frequencies*
A	$N_1\hat{p}_0$	$N_2\hat{p}_0$	X_1
$\bar{A}$	$N_1\hat{q}_0$	$N_2\hat{q}_0$	X_2
Total frequencies	N_1	N_2	N

identical, then a computational error has been made and should be corrected before proceeding. In any case, the test of $\theta = p_1 - p_2 = 0$, as computed by Pearson's classical test statistic, is equivalent to the Z test described in Section 13-5. This equivalence is now shown in the following example.

To show that two different test statistics give the same numerical value and, thereby, lead to the same conclusions, consider the data of Table 3-5, which lists the frequency of responses made by white students at two northern urban schools following a reorganization plan designed to integrate the schools The hypothesis to be tested is that student attitudes are unrelated to the school of attendance.

Stated in standard statistical notation, the hypothesis to be tested is $H_0\ p_1 = p_2 = p_0$ versus $H_1 \cdot H_0$ is false, where

$p_1 = P$ (liking school more|School A)

$p_2 = P$ (liking school more|School B)

$p_0 = P$ (liking school more)

Since H_1 is a nondirectional alternative and thereby requires a two-tailed test, the probability of a type I error for the Z test should be partitioned equally between the two tails of the normal distribution. For the X^2 form of the test, the only way that H_0 could be rejected would be if X^2 were too large. This would occur if the differences between the X_{ki} and $\hat{E}(X_{ki})$ were larger than could be expected on the basis of chance alone. Since the differences are squared, the probability of a type I error should be assigned only to the upper tail of the chi-square distribution, with one degree of freedom. If a one-tailed test should be desired, then the simplest procedure is to use Z and assign the probability of a type I error to the appropriate tail of the $N(0,1)$ distribution For this example, let $\alpha = .05$.

With Z as a test statistic, H_0 should be rejected if $Z < -1.96$ or if $Z > 1.96$. In this case, $\hat{p}_1 = \frac{189}{254} = .744$, $\hat{p}_2 = \frac{63}{121} = .521$, $\hat{p}_0 = \frac{252}{375} = .672$, and

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0\hat{q}_0/N_1 + \hat{p}_0\hat{q}_0/N_2}} = \frac{.744 - .521}{\sqrt{(.672)(.328)/254 + (.672)(.328)/121}}$$
$$= \frac{.223}{.052}$$
$$= 4.31$$

Since $Z = 4.31 > 1.96$, there is reason to believe that the attitudes of the white students are related to the school they attend. In fact, it appears that students at School A have a more favorable attitude toward school than the students at School B. Examination of the data in Table 3-3 shows that the same relationship holds for the Negro students in these two schools. For some reason, students at School B had unfavorable attitudes toward the school environment during the reorganization.

With X^2 as a test statistic, H_0 should be rejected if $X^2 > 3.84$. For this statistic, one must first set up the estimated expected frequency table. In setting up this table, it is advisable to carry one extra decimal value because rounding off errors may

Table 16-3. The estimated expected frequencies of Table 3-5

Response	*School A*	*School B*	*Total*
More	170.7	81.3	252
Less	83.3	39.7	123
Total	254	121	375

affect the value of X^2 With these estimated expected frequencies, as shown in Table 16-3,

$$X^2 = \sum_{k=1}^{2} \sum_{i=1}^{2} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$$

$$= \left[\frac{(189 - 170.7)^2}{170.7} + \frac{(65 - 83.3)^2}{83.3}\right] + \left[\frac{(63 - 81.3)^2}{81.3} + \frac{(58 - 39.7)^2}{39.7}\right]$$

$$= (1.96 + 4.02) + (4.12 + 8.44)$$

$$= 5.98 + 12.56$$

$$= 18.54 = (4.31)^2$$

As before, H_0 is rejected since $\chi_1^2(.95) = 3.84$. Since the differences between the X_{ki} and the corresponding $\hat{E}(X_{ki})$ are quite large relative to the $\hat{E}(X_{ki})$, the rejection of the hypothesis is not unexpected.

Since both test statistics lead to the same decision, it makes no difference which one is used for the two-sample problem. If a test is one-tailed, then it is easier to use Z.

16-2 THE K-SAMPLE BINOMIAL TEST

The extension of the K-sample binomial test of $H_0: p_1 = p_2 = \cdots = p_K = p_0$ versus the alternative that H_0 is false cannot be attained through the normal curve test, but X^2 can be extended to obtain a test of this hypothesis. If the data are tabulated

Table 16-4. Table of frequencies of the outcomes on K independent samples from binomial populations.

Outcome	*Sample 1*	*Sample 2*	⋯	*Sample K*	*Total frequency*
A	X_{11}	X_{21}	⋯	X_{K1}	X_1
$\bar{A}$	X_{12}	X_{22}	⋯	X_{K2}	X_2
Total frequency	N_1	N_2	⋯	N_K	N

as shown in Table 16-4, it is easy to see that the corresponding test statistic is given by

$$X^2 = \sum_{k=1}^{K} \sum_{i=1}^{2} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$$

As before, the assumptions are that $X_{11}, X_{12}, \ldots, X_{K1}$ are independent binomially distributed random variables. If these conditions are satisfied and if the expected frequencies are all greater than 5, then X^2 is approximately χ^2, with $\nu = (K-1)$ degrees of freedom. Since

$$\hat{p}_0 = \frac{X_{11} + X_{21} + \cdots + X_{K1}}{N}$$

$$= \frac{N_1 \hat{p}_1 + N_2 \hat{p}_2 + \cdots + N_K \hat{p}_K}{N}$$

it follows that

$$N\hat{p}_0 = (N_1 + N_2 + \cdots + N_K)\hat{p}_0 = N_1 \hat{p}_1 + N_2 \hat{p}_2 + \cdots + N_K \hat{p}_K$$

so that

$$N_1(\hat{p}_1 - \hat{p}_0) + N_2(\hat{p}_2 - \hat{p}_0) + \cdots + N_K(\hat{p}_K - \hat{p}_0) = 0$$

This relationship implies that the individual estimated deviations $(\hat{p}_k - \hat{p}_0)$ are not free to vary, since their weighted sum must add to 0. Thus, by the argument used to prove Theorems 6-3 or 10-1 about the sum of weighted deviations about a sample mean or expected value, it follows that the set of deviations possesses $\nu = (K-1)$ degrees of freedom. Therefore, X^2 is approximately χ^2 with $\nu = (K-1)$.

Another way to interpret ν is to note that K parameters $(p_1, p_2, \ldots, p_K)$ must be estimated from the data. However, these estimates are not independent, since they must satisfy the weighted deviation equation derived in the previous paragraph. In this sense, it is often stated that $(K-1)$ degrees of freedom are utilized to estimate the parameters of the model.

The use of this test statistic will be illustrated for the following data. In a mail survey of attitudes toward a school reorganization program, a random sample of adults was asked to respond to the following item:

> It has been suggested that grouping of students by ability sometimes leads to racial segregation in the classrooms. As a result, the school board would like to reduce the number of homogeneous groups and substitute heterogeneous grouping. Do you agree or disagree with this proposal?
>
> 1. I agree______ 2. I disagree______

The responses to this item are summarized in Table 16-5 by socioeconomic status of the respondents. On the basis of this evidence, can one conclude that attitude is

Table 16-5 Responses to the homogeneous-heterogeneous grouping item by socioeconomic status.

Response	*Socioeconomic status* LOW	MEDIUM	HIGH	*Total frequencies*
Agree	29	64	33	126
Disagree	47	164	135	346
Total frequencies	76	228	168	472

independent of social class? If it were, then it would be true that $p_L = p_M = p_H = p_0$, where

$$p_L = P\text{ (agree}|\text{low SES)}$$
$$p_M = P\text{ (agree}|\text{medium SES)}$$
$$p_H = P\text{ (agree}|\text{high SES)}$$
$$p_0 = P\text{ (agree)}$$

The sample estimates of these parameters are given by

$$\hat{p}_L = \tfrac{29}{76} = .3816$$
$$\hat{p}_M = \tfrac{64}{228} = .2807$$
$$\hat{p}_H = \tfrac{33}{168} = .1964$$
$$\hat{p}_0 = \frac{29 + 64 + 33}{76 + 228 + 168} = \frac{126}{472} = 2669$$

Table 16-6. Estimated expected frequencies of the data of Table 16-5.

Response	*Socioeconomic status* LOW	MEDIUM	HIGH	*Total frequencies*
Agree	20 3	60 9	44 8	126
Disagree	55 7	167 1	123 2	346
Total frequencies	76	228	168	472

In using X^2 as a test statistic, it is necessary to estimate the expected frequencies. These estimated expected frequencies are summarized in Table 16-6. For these data,

$$X^2 = \sum_{k=1}^{3}\sum_{t=1}^{2} \frac{[X_{kt} - \hat{E}(X_{kt})]^2}{\hat{E}(X_{kt})}$$

$$= \left[\frac{(29 - 20.3)^2}{20.3} + \frac{(47 - 55.7)^2}{55.7}\right] + \left[\frac{(64 - 60.9)^2}{60.9} + \frac{(164 - 167.1)^2}{167.1}\right] + \left[\frac{(33 - 44.8)^2}{44.8} + \frac{(135 - 123.2)^2}{123.2}\right]$$

$$= 5.0875 + .2153 + 4.2383$$

$$= 9.54$$

With $\alpha = .05$, the hypothesis of equal probability values is rejected since $\nu = K - 1 = 3 - 1 = 2$ and $\chi_2^2(.95) = 5.99$.

16-3 *POST HOC* PROCEDURES FOR THE *K*-SAMPLE BINOMIAL PROBLEM

For the data of the previous section there is reason to believe that attitudes differ across the three different social classes, but how they differ is still unknown. This is somewhat similar to the decision following the rejection of equal expected values in an analysis-of-variance study; one does not know why H_0 has been rejected. For the analysis-of-variance model, possible explanations could be found by direct application of Scheffé's method of multiple contrasts. Fortunately, Scheffé's theorem can be extended to cover the binomial case. As shown in Section 14-8, if ν_2 is large, then $\nu_1 F_{\nu_1,\nu_2}$ is approximately equal to $\chi^2_{\nu_1}$. Therefore, if $N = N_1 + N_2 + \cdots + N_K$ is large, so that ν_2 is large, the analogous Scheffé's coefficient is given by

$$S^* = \sqrt{\nu_1 F_{\nu_1,\infty}(1 - \alpha)} = \sqrt{\chi^2_{\nu_1}(1 - \alpha)}$$

The proof that this relationship holds and is valid for this case is beyond this book. As a result, the analog to Scheffé's theorem will be stated without proof and then the example of Section 16-2 will be used to show how the theorem can be employed to locate possible sources of rejection.

Theorem 16-1

The chi-square analog to Scheffé's theorem. In the limit the probability is $(1 - \alpha)$ that simultaneously for all linear contrasts of the form $\psi = a_1 p_1 + a_2 p_2 + \cdots + a_K p_K$, where $a_1 + a_2 + \cdots + a_K = 0$,

$$\hat{\psi} - \sqrt{\chi^2_{K-1}(1 - \alpha)}\,\mathrm{SE}_{\hat{\psi}} < \psi < \hat{\psi} + \sqrt{\chi^2_{K-1}(1 - \alpha)}\,\mathrm{SE}_{\hat{\psi}}$$

One of the interpretations that may be given to this theorem is that if $X^2 > \chi^2_{K-1}(1 - \alpha)$, then either the hypothesis of equal p values is not true or an event

of probability α has occurred. If $X^2 > \chi^2_{K-1}(1-\alpha)$, then there is at least one linear contrast of the parameters that is different from 0. And if $X^2 < \chi^2_{K-1}(1-\alpha)$, then all confidence intervals computed on a *post hoc* basis will be certain to include the value of 0. But note, if $X^2 > \chi^2_{K-1}(1-\alpha)$ and the decision is made to reject the hypothesis, then there exists at least one confidence interval that will not include 0 The interested researcher can look for it.

When $K=2$

$$\sqrt{\chi^2_{K-1}(1-\alpha)} = \sqrt{\chi^2_1(1-\alpha)} = \pm Z\left(\frac{\alpha}{2}\right)$$

so that the set of simultaneous confidence intervals reduces to the single familiar confidence interval for the difference between the two percentages, namely

$$(\hat{p}_1 - \hat{p}_2) - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1} + \frac{\hat{p}_2\hat{q}_2}{N_2}} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1} + \frac{\hat{p}_2\hat{q}_2}{N_2}}$$

For the example of Section 16-2,

$$\sqrt{\chi^2_{K-1}(1-\alpha)} = \sqrt{\chi^2_2(.95)} = \sqrt{5.99} = 2.45$$

The contrasts of interest and their numerical values are

$$\hat{\psi}_1 = \hat{p}_L - \hat{p}_M = .3816 - 2807 = .1009$$
$$\hat{\psi}_2 = \hat{p}_L - \hat{p}_H = .3816 - .1964 = .1852$$
$$\hat{\psi}_3 = \hat{p}_M - \hat{p}_H = .2807 - .1964 = .0843$$

Their estimated variances are given by

$$SE^2_{\hat{\psi}_1} = \frac{\hat{p}_L\hat{q}_L}{N_L} + \frac{\hat{p}_M\hat{q}_M}{N_M} = \frac{(.3816)(.6184)}{76} + \frac{(.2807)(.7193)}{228} = .00398$$

$$SE^2_{\hat{\psi}_2} = \frac{\hat{p}_L\hat{q}_L}{N_L} + \frac{\hat{p}_H\hat{q}_H}{N_H} = \frac{(.3816)(.6184)}{76} + \frac{(.1964)(.8036)}{168} = .00403$$

$$SE^2_{\hat{\psi}_3} = \frac{\hat{p}_M\hat{q}_M}{N_M} + \frac{\hat{p}_H\hat{q}_H}{N_H} = \frac{(.2807)(.7193)}{228} + \frac{(.1964)(.8036)}{168} = .00181$$

The $(1-\alpha) = .95$ set of simultaneous confidence intervals is given by

$$\psi_1 = .1009 \pm 2.45\sqrt{.00398} = .1009 \pm .1545$$
$$\psi_2 = .1852 \pm 2.45\sqrt{.00403} = .1852 \pm .1551$$
$$\psi^2 = .0843 \pm 2.45\sqrt{.00181} = .0845 \pm .1041$$

The only confidence interval that does not include 0 is ψ_2. Thus, it is concluded that a difference in attitude exists between the adults in the low and high socioeconomic classes.

After-the-fact inspection of the data suggests that the medium socioeconomic status group has attitudes similar to the adults in the high group, or that p_H and p_M may be equal and jointly different from p_L. This after-the-fact, or *post hoc*, hypothesis can be tested by the following complex contrast:

$$\psi_4 = \frac{N_M}{N_M + N_H} p_M + \frac{N_H}{N_M + N_H} p_H - p_L$$

If the hypothesis of no difference is true, then the expected value of this contrast should be 0 and a confidence interval about the estimated value should include 0. For these data,

$$\begin{aligned}
\hat{\psi}_1 &= \frac{228}{228 + 168}\hat{p}_M + \frac{168}{228 + 168}\hat{p}_H - \hat{p}_L \\
&= \left(\tfrac{228}{396}\right)\left(\tfrac{64}{228}\right) + \left(\tfrac{168}{396}\right)\left(\tfrac{33}{168}\right) - \tfrac{29}{76} \\
&= \frac{64 + 33}{396} - \frac{29}{76} \\
&= \tfrac{97}{396} - \tfrac{29}{76} \\
&= .2449 - .3816 \\
&= -.1367
\end{aligned}$$

$$\mathrm{SE}^2_{\hat{\psi}_4} = \left(\tfrac{228}{396}\right)^2 \mathrm{SE}^2_{\hat{p}_M} + \left(\tfrac{168}{396}\right)^2 \mathrm{SE}^2_{\hat{p}_H} + \mathrm{SE}^2_{\hat{p}_L} = .003567$$

$$\begin{aligned}
\psi_4 &= \hat{\psi}_4 \pm \sqrt{\chi^2_{K-1}(1-\alpha)}\sqrt{\mathrm{SE}^2_{\hat{\psi}_4}} \\
&= -.1367 \pm 2.45\sqrt{.003567} \\
&= -.1367 \pm .1463
\end{aligned}$$

Since 0 is included in the interval, the *post hoc* hypothesis is not supported.

16-4 EXTENSION OF THE *K*-SAMPLE BINOMIAL PROBLEM TO THE *K*-SAMPLE CHI-SQUARE TEST OF HOMOGENEITY

An important extension of the K-sample binomial problem to the case in which the individual populations are partitioned into more than two mutually exclusive and exhaustive subclasses is available through the extension of the chi-square statistic of Karl Pearson. This test is referred to in the literature as the chi-square test of homogeneity.

For this test, consider K populations from which random samples of size N_1, $N_2, \ldots, N_K$ are selected. After the samples are obtained, let each individual element be assigned to one and only one of the following set of mutually exclusive and exhaustive subsets $A_1, A_2, \ldots, A_I$, as shown in Table 16-7. When $I = 2$, the associated probability distributions are binomial. When $I = 3$, the distributions are trinomial, and when I is larger, the probability distributions are said to be multinomial. The

Table 16-7. Table of frequencies for the outcomes on K independent samples from multinomial populations.

Outcome	*Sample 1*	*Sample 2*	$\cdot$	*Sample K*	*Total frequencies*
A_1	X_{11}	X_{21}		X_{K1}	X_1
A_2	X_{12}	X_{22}	$\cdot$	X_{K2}	X_2
				$\cdot$	
	$\cdot$	$\cdot$	$\cdot\ \cdot$	$\cdot$	
$\cdot$	$\cdot$		$\cdot$	$\cdot$	
A_I	X_{1I}	X_{2I}	$\cdot$	X_{KI}	X_I
Total frequencies	N_1	N_2	$\cdots$	N_K	N

hypothesis to be tested is that the multinomial probability distributions for each of the populations are identical or homogeneous. Stated in statistical form, H_0 is given by

$$H_0\colon \begin{bmatrix} P(A_1|1) \\ P(A_2|1) \\ \vdots \\ P(A_I|1) \end{bmatrix} = \begin{bmatrix} P(A_1|2) \\ P(A_2|2) \\ \vdots \\ P(A_I|2) \end{bmatrix} = \cdots = \begin{bmatrix} P(A_1|K) \\ P(A_2|K) \\ \vdots \\ P(A_I|K) \end{bmatrix} = \begin{bmatrix} P(A_1) \\ P(A_2) \\ \\ P(A_I) \end{bmatrix}$$

and the alternative hypothesis is that H_0 is false. In this form it is seen that the alternative hypothesis indicates that this test is really a catchall or omnibus test. Anything that will make H_0 false is included in the set of alternative hypotheses. Thus, if H_0 is rejected, about the only way that one can locate the sources of variability is to use the *post hoc* method of the previous section. In this case, the method will prove to be cumbersome and time-consuming in performance. For that reason it is not illustrated.

The test statistic for this test is given by

$$X^2 = \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$$

which, when H_0 is true, has an approximate chi-square distribution with $\nu = (K-1)(I-1)$. While it is impossible to prove with the materials developed here that X^2 is approximately χ^2, it is easy to show that the degrees of freedom for the test are indeed given by $\nu = (K-1)(I-1)$. Since the estimated expected frequencies of sample 1 must add to N_1, it follows that the number of degrees of freedom available for the estimated frequencies of sample 1 is given by $(I-1)$. Thus, it appears that the total number of degrees of freedom available for the test is $K(I-1)$. But

since $p_1, p_2, \ldots, p_K$ must be estimated from the data, and since these estimates must satisfy the relation $N_1(\hat{p}_1 - \hat{p}_0) + N_2(\hat{p}_2 - \hat{p}_0) + \cdots + N_K(\hat{p}_K - \hat{p}_0) = 0$, it follows that $(I-1)$ degrees of freedom must be subtracted from $K(I-1)$. Thus $\nu = K(I-1) - (I-1) = (I-1)(K-1)$. The use of this test is illustrated in the following example.

As part of a study in which an attempt was made to determine if the attitudes of the members of a middle-class PTA could be swayed to approve student demonstrations, 173 parents were assigned to four different experimental conditions. The subjects assigned to condition 1 were shown two highly edited television films that placed most of the blame on student agitators, communists, and other revolutionaries. Subjects in condition 2 were shown one such film. Subjects in condition 3 were shown one film that gave a poor image of police enforcement. Subjects in condition 4 were shown two such highly edited films. One week before the film showing, a 10-item attitude questionnaire was given to each parent. Immediately following the film showing, the same test was given to each parent; difference scores were computed and classified as follows:

1. Attitude changed in a negative direction toward opposing student demonstrations.
2. Attitude remained the same.
3. Attitude changed in a positive direction toward supporting student demonstrations.

If the change score was an element of the set $\{+3, +2, +1, 0, -1, -2, -3\}$, it was said that no change occurred. Scores larger than $+3$ corresponded to a positive change while scores smaller than -3 corresponded to a negative change. The changes in attitude for 150 parents for whom complete data were available are summarized in Table 16-8. Even though the manipulated variable (kinds and number of films)

Table 16-8. Change in attitude for the four conditions of the study in which an attempt was made to change PTA members' attitudes toward student demonstrations.

Change in attitude	*Two anti-police films*	*One anti-police film*	*One anti-student film*	*Two anti-student films*	*Total*
Positive	28	20	5	0	53
Same	6	9	17	10	42
Negative	1	9	18	27	55
Total	35	38	40	37	150

and the criterion variable (change scores) are defined on an ordered scale, it *should not be assumed* that this test requires that the sets defining the variables be ordered.

In general, they are not. Very frequently, the categories are similar to the following partition used for race· (Caucasian, Oriental, Negro, other). Without doubt, this set cannot be ordered.

The experimental hypothesis of this example is that the way parents responded to student demonstrations was influenced by the kinds of films they were shown. At one level it seems reasonable to assume that parents who were shown antistudent films would show the greatest degree of annoyance with demonstrations, while at the other extreme it could be assumed that parents shown police brutality would side with the students and their causes. Thus, an ordered alternative on the parameters is quite meaningful Furthermore, visual inspection of the data supports this alternative.

It is worth emphasizing that the hypothesis to be tested is that the four trinomial probability distributions generated by the responses to the questionnaires are identical or homogeneous. Let C_1, C_2, C_3, and C_4 refer to the film conditions; then H_0 is

$$H_0\colon \begin{bmatrix} P(A_1|C_1) \\ P(A_2|C_1) \\ P(A_3|C_1) \end{bmatrix} = \begin{bmatrix} P(A_1|C_2) \\ P(A_2|C_2) \\ P(A_3|C_2) \end{bmatrix} = \begin{bmatrix} P(A_1|C_3) \\ P(A_2|C_3) \\ P(A_3|C_3) \end{bmatrix} = \begin{bmatrix} P(A_1|C_4) \\ P(A_2|C_4) \\ P(A_3|C_4) \end{bmatrix} = \begin{bmatrix} P(A_1) \\ P(A_2) \\ P(A_3) \end{bmatrix}$$

Unfortunately, H_0 can only be tested against an omnibus alternative Anything that will deny the truth of the hypothesis under test is considered as an appropriate alternative. Even though it is reasonable to expect certain ordered relations among the parameters as one goes across the film conditions, it is impossible to test for them. For example, it is reasonable to expect that if H_0 is false, then the parameters should satisfy the following sets of inequalities:

$$P(A_1|C_1) > P(A_1|C_2) > P(A_1|C_3) > P(A_1|C_4)$$

or

$$P(A_3|C_1) < P(A_3|C_2) < P(A_3|C_3) < P(A_3|C_4)$$

Unfortunately, H_0 could be rejected even if these relationships did not hold. This is because this X^2 test is insensitive to orderings on the variables.

If H_0 is true, then the best estimates of the common trinomial parameter values are

$$\hat{p}_1 = \frac{28+20+5+0}{35+38+40+37} = \frac{53}{150} = .3533$$

$$\hat{p}_2 = \frac{6+9+17+10}{35+38+40+37} = \frac{42}{150} = .2800$$

$$\hat{p}_3 = \frac{1+9+18+27}{35+38+40+37} = \frac{55}{150} = .3667$$

The estimated expected frequencies are found by multiplying these relative frequencies by the four different sample sizes. The estimated expected frequencies are summarized in Table 16-9. Once the estimated expected frequencies are determined,

Table 16-9. The estimated expected frequencies for the data of Table 16-8.

Change in attitude	*Two anti-police films*	*One anti-police film*	*One anti-student film*	*Two anti-student films*	*Total*
Positive	12.4	13.4	14.1	13.1	53
Same	9 8	10.6	11.2	10.4	42
Negative	12 8	13.9	14.7	13.6	55
Total	35	38	40	37	150

one should inspect to see that all marginal frequencies equal those of the observed frequency table. Also, the estimated expected frequencies should be checked to see that they are greater than 5. There is some evidence that if all expected frequencies do not exceed 5, one can still proceed with the test if less than 20 percent of the cells have estimated expected frequencies less than 5. In this case, all are greater than 5 so that one can conduct the test with confidence. Then we have

$$X^2 = \sum_{k=1}^{4} \sum_{i=1}^{3} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$$

$$= \left[\frac{(28-12.4)^2}{12.4} + \frac{(6-9.8)^2}{9.8} + \frac{(1-12.8)^2}{12.8}\right]$$

$$+ \left[\frac{(20-13.4)^2}{13.4} + \frac{(9-10.6)^2}{10.6} + \frac{(9-13.9)^2}{13.9}\right]$$

$$+ \left[\frac{(5-14.1)^2}{14.1} + \frac{(17-11.2)^2}{11.2} + \frac{(18-14.7)^2}{14.7}\right]$$

$$+ \left[\frac{(0-13.1)^2}{13.1} + \frac{(10-10.4)^2}{10.4} + \frac{(27-13.6)^2}{13.6}\right]$$

$$= 73.13$$

Since $X^2 = 73.13 > \chi_6^2(.95) = 12.59$, there is reason to doubt the truth of H_0. One can say that the probability distributions for the four experimental conditions are not the same.

This can be easily noted by examining the three-dimensional line graph of Figure 16-1, which shows the four conditional distributions defined by the conditions of

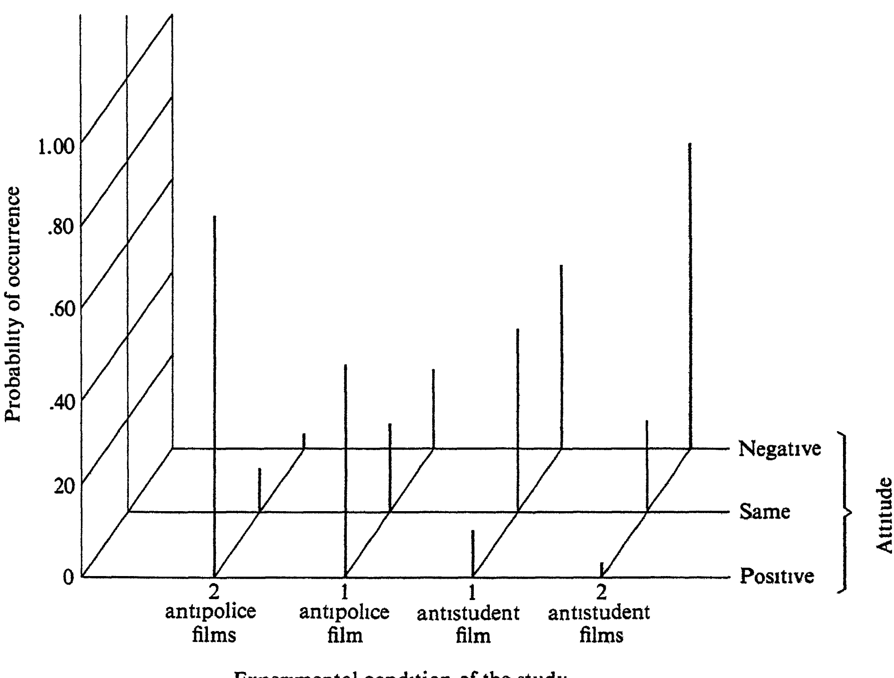
the study. In this case, the sum of the probabilities for any one of the conditional distributions is 1.

Since each category can be used to dichotomize the population, with the complementary set being everything that is not in the set of interest, it follows that binomial theory can be applied to each category, and *post hoc* analyses can be used to locate the sources of variation that perhaps lead to the rejection of the hypothesis. While this is not illustrated, visual inspection of the results suggests that the interpretation given earlier is a good one for these data and the population from which the data came.

16-5 SIMPLE COMPUTING FORMULA FOR CHI-SQUARE

Just as there are computing formulas for the analysis of variance, there exists a simple computing formula for the tests of this chapter. The reason it was not presented earlier is that it was hoped that an intuitive feeling for the test statistic could be developed by doing the computations according to the defining formulas of Pearson. The computing formula is best suited for most modern desk calculators. An experienced person can conduct most of the computation in five minutes or less The computing formula is stated as a consequence of the following theorem.

Figure 16-1. Conditional distributions of the four experimental conditions of the study



Theorem 16-2

The computing formula for X^2 is given by

$$X^2 = \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{X_{ki}^2}{\hat{E}(X_{ki})} - N$$

Proof

$$X^2 = \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{I} \left[\frac{X_{ki}^2}{\hat{E}(X_{ki})} - 2\frac{X_{ki}\hat{E}(X_{ki})}{\hat{E}(X_{ki})} + \frac{\hat{E}(X_{ki})^2}{\hat{E}(X_{ki})} \right]$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{X_{ki}^2}{\hat{E}(X_{ki})} - 2 \sum_{k=1}^{K} \sum_{i=1}^{I} X_{ki} + \sum_{k=1}^{K} \sum_{i=1}^{I} \hat{E}(X_{ki})$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{X_{ki}^2}{\hat{E}(X_{ki})} - 2N + N$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{X_{ki}^2}{\hat{E}(X_{ki})} - N$$

This completes the proof

In this form, if the individual estimated expected frequencies are almost equal to the observed frequencies, then the first term of the computing formula will be very close to N in numerical value, so that X^2 will be close to 0.

16-6 TABLE OF THE STATISTICAL TESTS AND CONFIDENCE INTERVALS FOR CHI-SQUARE TESTS OF HOMOGENEITY

The basic tests of this chapter are summarized in Table 16-10.

In some respects the hypothesis and *post hoc* confidence-interval procedure described in this chapter show a connection between the one-way analysis of variance for quantitative variables and the chi-square test of homogeneity for qualitative variables Because of this correspondence, the chi-square test of homogeneity can always be used as a substitute for the F test, provided that the basic data are categorized. An example of this substitution is provided in Exercises 16-5 and 16-7.

16-7 SUMMARY

In this chapter, a chi-square form of the test of the hypothesis $H_0 \cdot p_1 = p_2$ was presented. This test, algebraically equivalent to

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0 \hat{q}_0 / N_1 + \hat{p}_0 \hat{q}_0 / N_2}}$$

could be performed as

$$X^2 = \sum_{k=1}^{2} \sum_{i=1}^{2} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$$

where X_{ki} = observed frequency of the ith subset of population k and $\hat{E}(X_{ki})$ = estimated expected frequency of the ith subset of population k. If the observations are from two independent binomial universes such that the estimated expected frequencies are all greater than 5, it then follows that Z is approximately $N(0,1)$ and that X^2 is approximately χ_1^2. In the research literature, this test of hypothesis is most frequently performed in the classical Karl Pearson X^2 form.

For the test of $H_0: p_1 = p_2 = \cdots = p_K = p_0$, the most commonly used test statistic is

$$X^2 = \sum_{k=1}^{K} \sum_{i=1}^{2} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$$

which, when H_0 is true, has an approximate χ_{K-1}^2 distribution with $(1-\alpha)$ percent set of confidence intervals on the contrasts in the parameters given by

$$\hat{\psi} - \sqrt{\chi_{K-1}^2(1-\alpha)}\,\mathrm{SE}_{\hat{\psi}} < \psi < \hat{\psi} + \sqrt{\chi_{K-1}^2(1-\alpha)}\,\mathrm{SE}_{\hat{\psi}}$$

where

$$\hat{\psi} = a_1 \hat{p}_1 + a_2 \hat{p}_2 + \cdots + a_K \hat{p}_K$$

with

$$a_1 + a_2 + \cdots + a_K = 0$$

and

$$\mathrm{SE}_{\hat{\psi}}^2 = a_1^2\,\mathrm{SE}_{\hat{p}_1}^2 + a_2^2\,\mathrm{SE}_{\hat{p}_2}^2 + \cdots + a_K^2\,\mathrm{SE}_{\hat{p}_K}^2$$

When $I \geqslant 3$, the corresponding test of identical multinomial probability distributions is called the chi-square test of homogeneity. The test statistic for this test is

$$X^2 = \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$$

which, when H_0 is true, has a sampling distribution that is approximately chi-square with $\nu = (K-1)(I-1)$. The *post hoc* procedure for this model is identical to that for $I = 2$ and is obtained by separating the I sets into the set of interest along with the set consisting of that which remains.

While the conditions for this test require that each $\hat{E}(X_{ki}) > 5$, there is considerable evidence that the approximation of X^2 to χ^2 is quite good even if 20 percent of the $\hat{E}(X_{ki}) < 5$.

Table 16-10. Tests and confidence-interval procedures for K-sample qualitative model

Case	*Hypothesis*	*Test statistic*
15	H_0: $p_1 = p_2 = \cdots = p_K = p_0$ H_1: H_0 is false	$X^2 = \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$ where $\hat{E}(X_{ki}) = N_k \frac{X_i}{N}$
16	H_0: $\begin{bmatrix} P(A_1 \mid 1) \\ P(A_2 \mid 1) \\ \\ P(A_I \mid 1) \end{bmatrix} = \begin{bmatrix} P(A_1 \mid 2) \\ P(A_2 \mid 2) \\ \\ P(A_I \mid 2) \end{bmatrix} = \cdots = \begin{bmatrix} P(A_1 \mid K) \\ P(A_2 \mid K) \\ \\ P(A_I \mid K) \end{bmatrix} = \begin{bmatrix} P(A_1) \\ P(A_2) \\ \vdots \\ P(A_I) \end{bmatrix}$ H_1: H_0 is false	$X^2 = \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$ where $\hat{E}(X_{ki}) = N_k \frac{X_i}{N}$

While the correction for continuity was not mentioned in the performance of these tests, it is a wise idea to make the correction if the resulting numerical value of the test statistic is close to the critical value that leads to rejection of the hypothesis.

Table 16-11. Hypothetical data used to illustrate the correction for continuity in the chi-square test.

Outcome	*Sample 1*	*Sample 2*	*Total frequency*
A	20	30	50
$\bar{A}$	30	20	50
Total frequency	50	50	100

As an example, consider the uncorrected test for the set of data in Table 16-11. For these data,

$$X^2 = \frac{(20-25)^2}{25} + \frac{(30-25)^2}{25} + \frac{(30-25)^2}{25} + \frac{(20-25)^2}{25}$$

$$= 4$$

Confidence intervals	*Assumptions*
$p_{k_1} - p_{k_2} = (\hat{p}_{k_1} - \hat{p}_{k_2}) \pm \sqrt{\chi^2_{K-1}(1-\alpha)}\sqrt{\frac{\hat{p}_{k_1}\hat{q}_{k_1}}{N_{k_1}} + \frac{\hat{p}_{k_2}\hat{q}_{k_2}}{N_{k_2}}}$ $k_1, k_2 = 1, 2, \ldots, K$ $k_1 < k_2$	1. Independence between samples 2. Independence within samples 3. Binomial 4. Expected values greater than 5
$p_{k_1} - p_{k_2} = (\hat{p}_{k_1} - \hat{p}_{k_2}) \pm \sqrt{\chi^2_{K-1}(1-\alpha)}\sqrt{\frac{\hat{p}_{k_1}\hat{q}_{k_1}}{N_{k_1}} + \frac{\hat{p}_{k_2}\hat{q}_{k_2}}{N_{k_2}}}$ $k_1, k_2 = 1, 2, \ldots, K$ $k_1 < k_2$	1. Independence between samples 2. Independence within samples 3. Multinomial 4. Expected values greater than 5

which would lead to rejection of H_0 at $\alpha = .05$ To make the correction for continuity, one first adjusts the observed frequencies so as to make rejection more difficult. This correction is shown in Table 16-12. For these corrected data,

$$X^2 = \frac{(20.5-25)^2}{25} + \frac{(29.5-25)^2}{25} + \frac{(29.5-25)^2}{25} + \frac{(20.5-25)^2}{25}$$

$$= 3.24$$

which would not lead to rejection of H_0 at $\alpha = .05$. Since the approximation is better when the continuity correction has been made, the hypothesis that $p_1 = p_2$ should not be rejected on the basis of the evidence of Table 16-11.

Table 16-12. Data of Table 16-11 corrected for continuity.

Outcome	*Sample 1*	*Sample 2*	*Total frequency*
A	20.5	29.5	50
$\bar{A}$	29.5	20 5	50
Total frequency	50	50	100

Finally, the rule of thumb presented in Section 3-10 can now be ignored This rule stated: if the average of the positive deviations of $\hat{P}(A|B)$ and $\hat{P}(A|\bar{B})$ from $\hat{P}(A)$ does not exceed .10, A and B can be considered as being statistically independent. For the data of Table 16-11, $\hat{P}(A|B)=\frac{20}{50}=.4$, $\hat{P}(A|\bar{B})=\frac{30}{50}=$ 6, and $\hat{P}(A)=\frac{50}{100}=.5$, so that the average of the positive deviations is equal to $\frac{1}{2}[|.4-.5|+|.6-.5|]=.10$. If the corrected frequencies had been used, the average positive deviation would be equal to .09 so that one could operationally conclude independence of A and B. Since the appropriate statistical test has been developed, the rule of thumb of Section 3-10 can now be discarded. Furthermore, it should be noted that the rule of thumb was motivated for data similar to those of Table 16-11.

EXERCISES

***16-1.** Analyze the data of Table 3-2 on the basis of the methods presented in this chapter.

***16-2.** Analyze the data of Table 3-6 on the basis of the methods presented in this chapter.

***16-3.** As part of a class exercise in a beginning statistics course, 30 students were each asked to toss a six-sided die 10 times and report the results to their lab instructor. The results of the tabulation were as follows.

Outcome	1	2	3	4	5	6
Frequency	42	61	39	53	59	46

On the basis of this evidence, would you say that the die is fair? Note that this "goodness of fit" problem was not discussed in this chapter. However, the models of this chapter can be easily extended to this case by the Karl Pearson formula:

$$X^2=\sum_{k=1}^{6}\frac{[X_k-\hat{E}(X_k)]^2}{\hat{E}(X_k)}$$

***16-4.** According to the central limit theorem, the frequency table reported in Table 9-1 should represent a normally distributed variable with $E(X)=50$ and $\mathrm{Var}(X)=10^2/5=20$. To test this hypothesis, one can use the "goodness of fit" statistic of Exercise 16-3, by first determining from the table of the $N(0,1)$ distribution the probabilities of $\bar{X}$ being in the true intervals and then multiplying these probabilities by the sample size, $N=320$. Perform these operations and determine whether or not the distribution of means is, indeed, normal If the expected frequencies of the extreme intervals are less than 5, combine the intervals with their neighboring intervals before computing X^2.

***16-5.** If the expected frequencies are small, it is frequently necessary to combine intervals or cells to make the X^2 appropriate. For the data of Exercise 8-7,

(*a*) Delete the column of data for three semesters
(*b*) Combine the 0–999 frequencies with the 1,000–2,999 frequencies
(*c*) Combine the 10,000–19,999 frequencies with the 6,000–9,999 frequencies

and analyze the resulting table according to the methods of this chapter. What do you think of the arbitrary methods used to regroup these data? How would you treat the analysis so as not to lose data?

***16-6.** Analyze the data of Exercise 2-5 on the basis of the methods presented in this chapter.

***16-7.** In a study of the consumption of alcoholic beverages, 205 office employees of a large firm in San Francisco, California, volunteered to keep a monthly record of the amount of money spent on alcoholic beverages. The results were as shown in the table.

Amount spent	*Single men*	*Married men*	*Single women*	*Married women*	*Total*
Under $10	5	40	7	24	76
$10–$20	22	36	15	16	89
$20 or more	15	23	2	1	40
Total	42	99	24	41	205

Analyze these statistics in the light of the methods presented in this chapter.

***16-8.** For the data of Exercise 16-7, make the following contrasts.

(*a*) Single men versus married men
(*b*) Single men versus single women
(*c*) Single people versus married people

What do these findings suggest?

***16-9.** As part of an evaluation program in which three junior high schools were integrated, students were asked, "Did Negro and white students mix and talk to each other at your school this year?" Responses by schools are as shown.

Response	*School A*		*School B*		*Total*
	GRADE 7	GRADE 8	GRADE 7	GRADE 8	
More than last year	43	31	17	12	103
Same as last year	13	25	3	13	54
Less than last year	17	8	6	3	34
Total	73	64	26	28	191

What do these statistics suggest about the reorganization?

***16-10.** For the "More than last year" category of the data of Exercise 16-9, define a contrast for

(*a*) School *A* versus School *B*
(*b*) Grade 7 versus Grade 8
(*c*) (Grade 7 versus Grade 8 in School *A*) versus (Grade 7 versus Grade 8 in School *B*)

What do these contrasts measure? Are they different from 0?

17 INTRODUCTION TO CORRELATION THEORY IN THE QUALITATIVE MODEL

There's good news for you paunchy chaps who'd like to jog some of it off if you just had the time—

It doesn't take as much time as you may think, and may not have to be done as frequently.

This is the welcome, and surprising, conclusion of researchers at the University of California in Berkeley after 24 weeks of tests on two volunteer groups of joggers.

The tests were intended to pin down whether regular, endurance-type exercise is as good for us non-athletic types as has been claimed—and to find out scientifically if the runners are really on the right track

..In the beginning [Dr. Jack H.] Wilmore and Dr Joseph Royce put the volunteers through tests to determine capacity for work, heart rate and oxygen consumption. Blood samples were taken under the supervision of Dr Ernest Altekruse, a co-worker of [Dr. Kenneth H.] Cooper

Then the subjects ran around the track at least twice a week, for brief periods at first and gradually increasing the time and speed to $1\frac{1}{2}$ miles in 12 minutes for Group I and three miles in 24 minutes for Group II. It was assumed that the group running twice as long would show more training effect, the question was how much more.

...Blood pressures decreased 10 to 13 percent, and the resting heart rate decreased an average of 12 percent. Specific cases varied widely A flabby 37-year-old's resting pulse rate was 102 at the beginning of the tests and 75 at the end, another man's dropped from 94 to 57, an athlete already in nearly fit condition started the program with a resting pulse rate of 66 and dropped to 63. The average was 76 at the start and 66 after 12 weeks.

Oxygen uptake increased an average 10 percent, a big increase in terms of energy and work capacity. Efficiency of oxygen use increased 13 percent, maximum work capacity increased 16 percent.

But the shocker "There was no statistical difference in results between Group I, the 12 minute runners, and Group II, the 24 minute runners.". .

In *Jogging* by Jim Quint, courtesy of California Living, August 10, 1969, The San Francisco Examiner.

17-1 CHI-SQUARE TESTS OF HOMOGENEITY AND INDEPENDENCE—A COMPARISON

Chi-square tests of independence are frequently confused with chi-square tests of homogeneity, both by beginning students and by experienced behavioral research scientists. This almost universal confusion is easy to understand since the test statistics for both tests are identical. Furthermore, both test statistics have sampling distributions that are approximately chi-square with the same number of degrees of freedom. Finally, the hypotheses tested by both test statistics are very similar. For the two-sample chi-square test of homogeneity, the tested hypothesis is $H_0: P(A|B) = P(A|\bar{B})$, while in the one-sample chi-square test of independence, the tested hypothesis is H_0 $P(A \cap B) = P(A)P(B)$. As was seen in Section 3-10, these probability statements are interchangeable and both are valid when the two characteristics A and B are statistically independent.

While it is easy to develop the theory for the chi-square test of homogeneity with the methods introduced in this book, the same is not true for the chi-square tests of independence. The mathematical theory behind these tests is, in some respects, the most complicated of the procedures ever encountered in classical statistical theory. For this reason, no attempt will be made to develop these tests. Instead, a weak heuristic argument will be presented to justify their use. This argument is based upon the Karl Pearson statistic, which, in this case, has intuitive appeal.

17-2 CHI-SQUARE TEST OF INDEPENDENCE FOR A 2 × 2 CONTINGENCY TABLE

Consider a *single* random sample of predetermined size N selected from some universe or population of interest. On each element of the obtained sample let simultaneous measurements be made on two variables that are each dichotomous or binomial. Let the variables be denoted by Y_1 and Y_2. Since two variables are being observed, this is referred to as a bivariate problem, and the test is a bivariate test of hypothesis. Let the characteristics associated with Y_1 be denoted by A and $\bar{A}$, and let those associated with Y_2 be denoted by B and $\bar{B}$. Let X_{11} represent the number of elements of the sample that possess both properties A and B and let this notation be extended to the three remaining mutually exclusive intersecting subsets. Also, let $X_{11} + X_{12} = X_{1.}$ and let this dot notation be extended to the remaining totals. With this notation, the observed frequencies can be summarized in a table similar to that of Table 17-1, which is popularly referred to as a 2×2 *contingency table*. Though the notation is different, the table is identical to Table 3-4. This means that the discussion of statistical independence introduced in Chapter 3 is being completed by the presentation of the classical test of the hypothesis that two dichotomous characteristics are statistically independent.

The important property about Table 17-1 that makes it different from Table 16-1 is that the only number that is known prior to the experiment or collection of data is N, the sample size. The numbers in both margins of the table are random variables, as are the numbers in the body of the table. For the chi-square test of homogeneity, N_1 and N_2 are known in advance The remaining numbers in the body of the table

Table 17-1. The 2 × 2 contingency table for the joint classification of two binomial variables measured on one sample.

Variable Y_2	*Variable* Y_1 OUTCOME A	OUTCOME $\bar{A}$	*Total frequencies*
OUTCOME B	X_{11}	X_{21}	$X_{.1}$
OUTCOME $\bar{B}$	X_{12}	X_{22}	$X_{.2}$
Total frequencies	X_1	X_2	N

are random variables. Whereas the chi-square test of homogeneity tests the hypothesis that $P(A|B) = P(A|\bar{B})$, the hypothesis tested with the chi-square test of independence is that the variables Y_1 and Y_2 are statistically independent or that $P(A \cap B) = P(A)P(B)$. If this hypothesis is true, then it is also true that $P(A \cap \bar{B}) = P(A)P(\bar{B})$, $P(\bar{A} \cap B) = P(\bar{A})P(B)$, and $P(\bar{A} \cap \bar{B}) = P(\bar{A})P(\bar{B})$. If these probabilities were known, one could determine the expected frequencies since it would be true that $E(X_{11}) = NP(A)P(B)$, with corresponding expected frequencies for the remaining cells. Since these probabilities are unknown, one can estimate them from the observed data and then use them to estimate the cell frequencies under the assumption that the hypothesis of independence is true. This would produce an observed table of frequencies and a table of estimated expected frequencies for which one could compute a chi-square statistic in the Karl Pearson form. In this case, the statistic would be given by

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{[X_{ij} - \hat{E}(X_{ij})]^2}{\hat{E}(X_{ij})}$$

and the hypothesis of independence would be rejected if this statistic exceeded the $(1 - \alpha)$ percentile of the chi-square distribution with one degree of freedom.

In the Berkeley School District reorganization study discussed earlier, a study was made of attitudes toward the integration of the elementary and secondary schools. The statistics for this study are reported in Table 17-2. The estimates of the unknown parameters are given by

$$\hat{P}(A) = \tfrac{60}{212} \qquad \hat{P}(\bar{A}) = \tfrac{152}{212} \qquad \hat{P}(B) = \tfrac{103}{212} \qquad \text{and} \qquad \hat{P}(\bar{B}) = \tfrac{109}{212}$$

The estimated expected frequencies are given by

$$\hat{E}(X_{11}) = 212\,(\tfrac{60}{212})(\tfrac{103}{212}) = 29\ 15$$
$$\hat{E}(X_{21}) = 212\,(\tfrac{152}{212})(\tfrac{103}{212}) = 73.85$$
$$\hat{E}(X_{12}) = 212\,(\tfrac{60}{212})(\tfrac{109}{212}) = 30.85$$
$$\hat{E}(X_{22}) = 212\,(\tfrac{152}{212})(\tfrac{109}{212}) = 78.15$$

Table 17-2. The 2 × 2 contingency table for the joint responses to the questions concerning (1) the integration of elementary schools, and (2) the integration of secondary schools.

Question 2	*Question 1* AGREE	DISAGREE	*Total*
AGREE	48	55	103
DISAGREE	12	97	109
Total	60	152	212

Table 17-3. Estimated frequencies for the data of Table 17-2.

Question 2	*Question 1* AGREE	DISAGREE	*Total*
AGREE	29.15	73.85	103
DISAGREE	30 85	78 15	109
Total	60	152	212

These frequencies are summarized in Table 17-3, and the value of the test statistic is given by

$$X^2 = \frac{(48 - 29.15)^2}{29.15} + \frac{(55 - 73.85)^2}{73.85} + \frac{(12 - 30.85)^2}{30.85} + \frac{(97 - 78.15)^2}{78.15}$$
$$= 33.07$$

Since 33.07 > 3.84, there is reason to doubt the hypothesis of independence. Therefore, it is concluded that the attitudes toward the integration of the elementary and secondary schools are not independent. Individuals who oppose one proposition tend to oppose the other, while individuals who support one proposition tend to support the other. Visual inspection of the data of Table 17-2 suggests that the strength of the association is rather modest and not too strong since the number of individuals with different points of view on the two propositions is relatively large. If the association had been strong, one would have found most of the observations along the main diagonal, suggesting a consistent point of view toward the two propositions.

Since this is a one-sample test, the *post hoc* procedures based on Scheffé's theorem are not appropriate. The appropriate *post hoc* procedure that one can employ to further investigate the hypothesis is discussed in Section 17-5.

If the estimated expected frequencies are less than 5, then the approximation to χ_1^2 might not be adequate. If this should occur, one can use the hypergeometric probabilities of Table A-3 to test for independence The manner in which this table is used is identical to the procedure described in Section 5-9 for testing $H_0: P(A|B) = P(A|\bar{B})$.

17-3 THE CHI-SQUARE TEST OF INDEPENDENCE FOR $I \times J$ CONTINGENCY TABLES

The test of the previous section can be extended to cover the case in which the Y_1 variable is partitioned into I mutually exclusive and exhaustive subsets $\{A_1, A_2, \ldots, A_I\}$ and in which the Y_2 variable is partitioned into J mutually exclusive and ex-

Table 17-4. A typical $I \times J$ contingency table.

Variable Y_2	*Variable* Y_1 A_1	A_2	A_3	·	A_I	*Total*
B_1	X_{11}	X_{21}	X_{31}		X_{I1}	$X_{\cdot 1}$
B_2	X_{12}	X_{22}	X_{32}	···	X_{I2}	$X_{\cdot 2}$
B_3	X_{13}	X_{23}	X_{33}	·	X_{I3}	$X_{\cdot 3}$
·		·	·	··		·
·	·	·	·	···	·	
·	·	·	·	··	·	·
B_J	X_{1J}	X_{2J}	X_{3J}	··	X_{IJ}	$X_{\cdot J}$
Total	$X_{1\cdot}$	$X_{2\cdot}$	$X_{3\cdot}$	···	$X_{I\cdot}$	N

haustive subsets $\{B_1, B_2, \ldots, B_J\}$. If the data are as shown in Table 17-4, the appropriate test statistic for testing statistical independence is given by

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{[X_{ij} - \hat{E}(X_{ij})]^2}{\hat{E}(X_{ij})}$$

where $\hat{E}(X_{ij}) = N\hat{P}(A_i)\hat{P}(B_j)$.

In this case, the distribution of X^2 is approximately chi-square with $(I-1)(J-1)$ degrees of freedom. Since $N = X_{11} + X_{12} + X_{13} + \cdots + X_{IJ}$, the number of independent cell frequencies is equal to $IJ - 1$. The number of independent parameters to be estimated for the A classification is $(I-1)$ and the number of independent parameters to be estimated for the B classification is $(J-1)$, so that the total number of degrees of freedom for the test statistic is given by $\nu = (IJ-1) - (I-1) - (J-1) = (I-1)(J-1)$. The use of this test will be illustrated by an example.

In a study conducted in a small city in western New York in 1967, 67 parents of medium family income level were asked to respond to an attitude questionnaire in which they were quizzed about the local school administration, certain school programs, and general school problems. The questionnaires were administered to the parents prior to the institution of an experimental reading program in which their children were to participate. During the experimental period parents were told about the program, they were informed about how well their children were doing in the program, and they were given a general education about the school and its objectives. After the completion of the program, the parents were given the same questionnaire and were asked to report their attitudes on the same questions asked at the beginnings of the study. On both tests high scores represented positive attitudes toward the administration, the program, and the schools. In addition to the many questions concerning the change in parents' attitudes, one would like to know whether there was a consistent relationship between the two sets of responses. That is, did parents with favorable responses maintain their positive attitudes and did those with negative attitudes retain theirs? As a research question, one would like to know whether the responses on the two testings are statistically independent. If the responses are not statistically independent, they are said to be correlated.

The classical correlation procedure for testing this hypothesis is discussed in Chapter 18. Since the assumptions for the use of this standard or classical procedure were not valid in this case, it was decided to trichotomize scores on the pretest and posttests to see whether the scores on the two tests maintained their same relative positions. The joint frequency table for pretest and posttest scores is as shown in Table 17-5. Since the observed frequency distributions were partitioned at the

Table 17-5. The 3 × 3 contingency table for the joint responses to the questionnaire on pretests and posttests of the 67 adult citizens of the study.

Posttest	*Pretest*			*Total*
	LOWER THIRD	MIDDLE THIRD	UPPER THIRD	
UPPER THIRD	7	5	11	23
MIDDLE THIRD	3	14	4	21
LOWER THIRD	11	7	5	23
Total	21	26	20	67

$33\frac{1}{3}$ and $67\frac{2}{3}$ percentile values, one would expect the marginal totals of the table to be equal to one another. Normally, this would be the case. However, in this study, there were an unusual number of observed values tied at the dividing percentile values. For each of these tied values, a coin was flipped to determine which adjoining cell a questionable observation should be assigned. For this reason, the marginal frequencies are unequal. In any case, the hypothesis tested is that the

responses to the two questionnaire testings are uncorrelated, unrelated, or statistically independent. As a statistical hypothesis, this reduces to $H_0: P(A_i \cap B_j) = P(A_i)P(B_j)$, where $i = 1, 2, 3$ corresponds to lower, middle, and upper thirds on the Y_1 variable and where $j = 3, 2, 1$ corresponds to the lower, middle, and upper thirds on the Y_2 variable. The alternative hypothesis is H_1 H_0 is false.

The statistical association between pretest and posttest scores can be seen easily by examining the three-dimensional line graph in Figure 17-1 for the joint distribution of the two variables. As is quite apparent, the largest probabilities are related to the diagonal elements with $11 + 14 + 11 = 36$ out of 67 parents expressing consistent points of view. Unlike the data of Figure 16-1, the sum of the individual probabilities adds to 1, since there is only one sample involved in the study.

The test statistic for this test is

$$X^2 = \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{[X_{ij} - \hat{E}(X_{ij})]^2}{\hat{E}(X_{ij})}$$

Since $\nu = (I - 1)(J - 1) = (3 - 1)(3 - 1) = 4$, the hypothesis of statistical independence should be rejected if $X^2 > \chi_4^2(.95) = 9.48$. The estimated expected frequencies under the hypothesis of statistical independence required for the test are summarized in Table 17-6.

Figure 17-1. Joint distribution of responses on the pretests and posttests

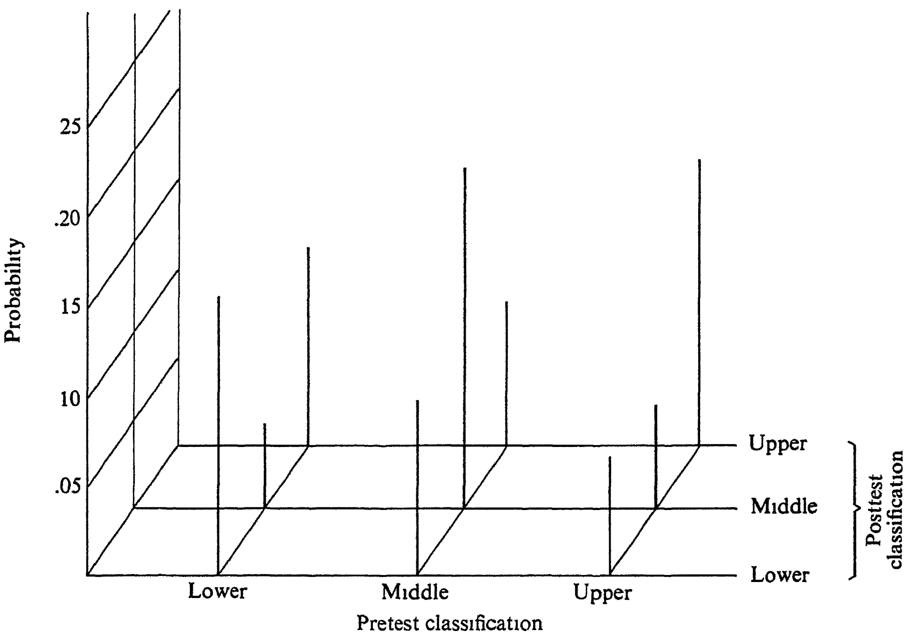


Table 17-6. The estimated expected frequencies for the data of Table 17-5.

Posttest	*Pretest*			*Total*
	LOWER THIRD	MIDDLE THIRD	UPPER THIRD	
UPPER THIRD	7.21	8.93	6.87	23
MIDDLE THIRD	6.58	8.15	6.27	21
LOWER THIRD	7.21	8.93	6.87	23
Total	21	26	20	67

The value of the test statistic is given by $X^2 = 14.11$. Since $14.11 > 9.48$, H_0 is rejected, and it is concluded that the responses to the questionnaire on two repeated testings are not independent. Visual inspection of the data suggests that parents with negative attitudes maintained their negative attitudes while parents with positive attitudes continued to have positive attitudes. The *post hoc* procedure for this test is discussed in Section 17-6.

As with most of the other statistical tests discussed in this book, an extremely strong and important assumption that must be made for the distribution of X^2 to be adequately described by χ^2 with $\nu = (I-1)(J-1)$ is that the observations be independent. If both parents of some of the families participated in the study, then the violation of the assumption is almost assured, since husbands and wives tend to possess similar attitudes. One other condition that might invalidate the assumption of statistical independence between parents of different families is verbal discussion of the pretest questions before the administration of the posttest. A persuasive opinionated parent can influence a great number of uncommitted parents who might never have thought of the problems or issues that a school administrator must contend with in doing his job successfully. If the issues are controversial, such as the teaching of sex education in elementary schools, then the effects upon the assumption of statistical independence can be quite pronounced. As mentioned earlier, and repeated here, the assumption of independence is vital to hypothesis testing and parameter estimation and its major guarantee is in the hands of a competent researcher. It is not something that appears automatically in a study. It is something that the researcher must *deliberately* build into his investigation. If it does not exist, then one should not persist in doing statistical tests. It may be acceptable to fool one's self, but it is dangerous and immoral to fool other researchers and individuals who use published research results to advance their own scientific studies or to suggest social action programs in the community.

17-4 STRENGTH OF ASSOCIATION IN THE QUALITATIVE CORRELATION MODEL

In Section 13-6, the connection between hypothesis testing for equal parameter values and testing for equal conditional probabilities was discussed. It was seen that denial of H_0: $\mu_1 = \mu_2$ or H_0: $p_1 = p_2$ was equivalent to concluding that the

associated probability distributions were unequal. For analysis-of-variance designs, it was seen in Section 15-9 that the corresponding situation existed when $H_0 . \mu_1 = \mu_2 = \cdots = \mu_K = \mu_0$ was rejected on the basis of an F test, but in this case it was noted that one could measure the degree or strength of association by means of a statistic $\hat{\omega}^2$, which represented the percentage of variation in the criterion variable that could be attributed to the manipulated variable. For qualitative correlational studies, a similar measure is available for describing the strength of association between two qualitative variables. This measure is called the phi coefficient and is denoted by the Greek letter $\hat{\phi}$. This measure is directly related to Karl Pearson's X^2 statistic. When $X^2 = 0$, it follows that Y_1 and Y_2 are statistically independent. This also means that Y_1 and Y_2 are unrelated, uncorrelated, or unassociated. When this occurs, $\hat{\phi}^2 = 0$ or some number very close to 0. On the other hand, when X^2 takes on its maximum possible value, it follows that Y_1 and Y_2 are perfectly related. When this occurs, $\hat{\phi}^2 = 1$, or some number very close to 1 Thus, the range in $\hat{\phi}^2$ of 0 to 1 is similar to the range in $\hat{\omega}^2$ as a measure of association or strength of association. This measure is derived in the following paragraphs.

In Section 6-10, the variance of a discrete random variable Y_1, which can take on the values $\{y_{11}, y_{12}, \ldots, y_{1I}\}$, was defined as

$$\sigma_1^2 = \text{Var}(Y_1) = \sum_{i=1}^{I} (y_{1i} - \mu_1)^2 P[Y_1 = y_{1i}]$$

where $E(Y_1) = \mu_1$; σ_1^2 can also be written as

$$\sigma_{11} = \sum_{i=1}^{I} (y_{1i} - \mu_1)(y_{1i} - \mu_1) P[Y_1 = y_{1i}]$$

An unbiased estimate of σ_1^2, based on a sample of size N, is given by S_1^2 ,which can be written as

$$S_{11} = \sum_{i=1}^{N} \frac{(y_{1i} - \bar{Y}_1)(y_{1i} - \bar{Y}_1)}{N - 1}$$

with the computing formula given by

$$S_{11} = \frac{N \sum_{i=1}^{N} y_{1i}^2 - \left(\sum_{i=1}^{N} y_{1i} \right)^2}{N(N-1)}$$

In an analogous fashion, the covariance of two discrete random variables Y_1 and Y_2, which can take on the values $\{y_{11}, y_{12}, \ldots, y_{1I}\}$ and $\{y_{21}, y_{22}, \ldots, y_{2J}\}$, respectively, can be defined as

$$\sigma_{12} = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{1i} - \mu_1)(y_{2j} - \mu_2) P[Y_1 = y_{1i} \cap Y_2 = y_{2i}]$$

While a proof is not presented, one can show that an unbiased estimate of σ_{12}, based upon a sample of size N, is given by

$$S_{12} = \sum_{i=1}^{N} \frac{(y_{1i} - \bar{Y}_1)(y_{2i} - \bar{Y}_2)}{N-1}$$

with computing formula given by

$$S_{12} = \frac{N \sum_{i=1}^{N} y_{1i} y_{2i} - \left(\sum_{i=1}^{N} y_{1i}\right)\left(\sum_{i=1}^{N} y_{2i}\right)}{N(N-1)}$$

While a detailed discussion of covariance is presented in Chapter 18, it is sufficient to state at this point that the covariance of two variables measures the degree to which the variables vary together. If the variables show no tendency to vary together, they are said to be statistically independent and their covariance is 0, as is shown in Theorem 17-1. However, if Y_1 and Y_2 are not independent, then their covariance is different from 0.

Theorem 17-1

If Y_1 and Y_2 are statistically independent, then $\text{Cov}(Y_1, Y_2) = 0$.

Proof By definition,

$$\text{Cov}(Y_1, Y_2) = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{1i} - \mu_1)(y_{2i} - \mu_2) P[Y_1 = y_{1i} \cap Y_2 = y_{2i}]$$

If Y_1 and Y_2 are statistically independent, then

$$P[Y_1 = y_{1i} \cap Y_2 = y_{2i}] = P[Y_1 = y_{1i}] P[Y_2 = y_{2i}]$$

so that

$$\text{Cov}(Y_1, Y_2) = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{1i} - \mu_1)(y_{2i} - \mu_2) P[Y_1 = y_{1i}] P[Y_2 = y_{2i}]$$

$$= \left(\sum_{i=1}^{I} (y_{1i} - \mu_1) P[Y_1 = y_{1i}]\right)\left(\sum_{j=1}^{J} (y_{2i} - \mu_2) P[Y_2 = y_{2j}]\right)$$

By Theorem 6-3, the algebraic quantities in each of the pairs of brackets is equal to 0, since each is the weighted sum of the deviations about its corresponding expected value Thus, $\text{Cov}(Y_1, Y_2) = 0$. This completes the proof.

The practical consequence of this theorem is that $\text{Cov}(Y_1, Y_2) = 0$ if Y_1 and Y_2 are statistically independent and $\text{Cov}(Y_1, Y_2) \neq 0$ if Y_1 and Y_2 are not statistically independent. While one might be tempted to conclude that if $\text{Cov}(Y_1, Y_2) = 0$, it must follow that Y_1 and Y_2 are statistically independent, one is advised not to make this error, for it is possible to find examples from behavioral studies in which the $\text{Cov}(Y_1, Y_2) = 0$ and yet find that the variables are not statistically independent.

Examples of this nature are presented in Chapter 18. In any case, it is true that the $\mathrm{Cov}(Y_1, Y_2)$ can be used as a measure of the strength of the association between two variables. Large values of $\mathrm{Cov}(Y_1, Y_2)$ correspond to strong associations, while values close to 0 *generally* correspond to statistical independence of the two variables.

Even though the covariance is a valid measure of association, it is customary to transform it to a scale such that no correlation is indicated by 0 while perfect correlation is indicated by ± 1, with the algebraic sign being indicative of the direction of the association; a plus sign denotes that Y_1 and Y_2 increase together while a negative sign means that one variable increases as the other variable decreases The transformation to the scale of ± 1 with 0 representing no relationship is attainable by simply dividing the $\mathrm{Cov}(Y_1, Y_2)$ by $\sigma_1 \sigma_2$. This transformed measure is called the *correlation coefficient*. For contingency tables it is denoted by the Greek letter ϕ (phi) and is defined as

$$\phi = \frac{\mathrm{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2}$$

A sample estimate of ϕ is given by

$$\hat{\phi} = \frac{S_{12}}{S_1 S_2}$$

This estimate is not unbiased, since it can be shown that $E(\hat{\phi}) \neq \phi$. The estimated correlation value is called the sample phi coefficient and is the appropriate measure of correlation for data reported in an $I \times J$ contingency table.

17-5 THE CORRELATION COEFFICIENT FOR A 2 X 2 CONTINGENCY TABLE

Any qualitative dichotomous variable can be quantified by associating a $+1$ if an A is observed and a 0 if an $\bar{A}$ is observed.

For a sample of size N, as reported in Table 17-1, one finds that

$$\sum_{i=0}^{1} y_{1i} = (1)X_1 \ + (0)X_{2.} = X_1$$

$$\sum_{i=0}^{1} y_{1i}^2 = (1)^2 X_1 \ + (0)^2 X_{2.} = X_1$$

so that

$$S_1^2 = S_{11} = \frac{N(X_1\ ) - X_1^2}{N(N-1)}$$

For the Y_2 variable, the corresponding quantification is achieved by assigning a +1 when a B is observed while a 0 is assigned if a $\bar{B}$ is observed Thus, for the N trials reported in Table 17-1,

$$\sum_{j=0}^{1} y_{2j} = (1)X_{.1} + (0)X_{.2} = X_{.1}$$

$$\sum_{j=0}^{1} y_{2j}^2 = (1)^2 X_{.1} + (0)^2 X_{.2} = X_{.1}$$

so that

$$S_2^2 = S_{22} = \frac{N(X_{.1}) - X_{.1}^2}{N(N-1)}$$

Under the {0,1} quantification of the two sets of classes,

$$\sum_{i=0}^{1} \sum_{j=0}^{1} y_{1i} y_{2j} = (1)(1)X_{11} + (1)(0)X_{12} + (0)(1)X_{21} + (0)(0)X_{22}$$
$$= X_{11}$$

so that

$$S_{12} = \frac{N(X_{11}) - (X_{1.})(X_{.1})}{N(N-1)}$$

If we substitute the corresponding algebraic quantities into

$$\hat{\phi} = \frac{S_{12}}{S_1 S_2}$$

we have

$$\hat{\phi} = \frac{(NX_{11} - X_{1.} X_{.1})/N(N-1)}{\sqrt{(NX_{1.} - X_{1.}^2)/N(N-1)}\,\sqrt{(NX_{.1} - X_{.1}^2)/N(N-1)}}$$
$$= \frac{NX_{11} - X_{1.} X_{.1}}{\sqrt{NX_{1.} - X_{1.}^2}\,\sqrt{NX_{.1} - X_{.1}^2}}$$

which is referred to as the computing formula for $\hat{\phi}$. Note that this formula is dependent upon only four figures, reported in Table 17-1 These figures are:

1. N: the total sample size
2. $X_{1.}$: the number of observations that have property A
3. $X_{.1}$: the number of observations that have property B
4. X_{11}: the number of observations that have both property A and property B

For the data of Table 17-2,

$$\hat{\phi} = \frac{212(48) - (60)(103)}{\sqrt{212(60) - (60)^2}\sqrt{212(103) - (103)^2}}$$

$$= .3949$$

For behavioral data, this can be considered as a moderate degree of association but not of much practical utility. Since the value of the test statistic $X^2 = 33.07$ leads to the conclusion that Y_1 and Y_2 were not statistically independent, $\hat{\phi} = .3949$ represents a correlation that is statistically different from 0. This suggests that the test of the hypothesis H_0 $\phi = 0$ versus the alternative H_1 $\phi \neq 0$ is identical to the test of the hypothesis H_0 $P(A \cap B) = P(A)P(B)$ versus the alternative H_1 H_0 is false, and indeed this is the case.

Since the computation of $\hat{\phi}$ may appear difficult, there is an easier way, provided that X^2 has been previously determined. This simplification is based upon a simple algebraic relationship between $\hat{\phi}$ and X^2. This relationship is shown in Theorem 17-2, which is stated without proof.

Theorem 17-2

The $\hat{\phi}$ coefficient from a 2×2 contingency table is given by

$$\hat{\phi} = \sqrt{\frac{X^2}{N}}$$

where X^2 is the Karl Pearson statistic for the table and N is the sample size.

For the data of Table 17-2, $N = 212$ and $X^2 = 33.07$, so that

$$\hat{\phi} = \sqrt{\frac{33.07}{212}} = .3949$$

which is identical to the value determined from the computing formula.

17-6 THE CORRELATION COEFFICIENT FOR AN $I \times J$ CONTINGENCY TABLE

For an $I \times J$ contingency table the defining formula for $\hat{\phi}$ is very complex. Fortunately, there is a simple computational formula that reduces to the formula of Theorem 17-2 when $I = J = 2$ This formula is stated without derivation and proof in Theorem 17-3.

Theorem 17-3

The $\hat{\phi}$ coefficient for an $I \times J$ contingency table is given by

$$\hat{\phi} = \sqrt{\frac{X^2}{NM}}$$

where X^2 is the Karl Pearson statistic for the table, N is the sample size, and M is the minimum of $(I - 1)$ and $(J - 1)$.

For the data of Table 17-4, $X^2 = 14\ 11$, $N = 67$, $(I - 1) = (3 - 1) = 2$, $(J - 1) = (3 - 1) = 2$, and $M = 2$, so that

$$\hat{\phi} = \sqrt{\frac{14\ 11}{67(2)}} = \ 3244$$

This correlation represents a moderate degree of association which, again, is not of much practical utility

Finally, it should be noted that $\hat{\phi}^2$ is not influenced by permuting rows or columns in an $I \times J$ contingency table. Because of this, the categories of Y_1 and Y_2 need not be on an ordered scale. For example, Y_1 may be partitioned according to father's occupation {professional, business manager, blue collar, other}, while Y_2 may be partitioned according to father's religion {Protestant, Jew, Catholic, other} Thus, if one tests the hypothesis that father's occupation is independent of his religion and if the hypothesis is rejected, one knows that $\phi^2 \neq 0$; however, a simple interpretation of ϕ^2 may not be possible In any case, if the categories of Y_1 or Y_2 are shifted, the estimate of ϕ^2 will not change, its value is independent of the ordering used in the tabulation of cell frequencies

17-7 THE EFFECT OF SAMPLE SIZE ON X^2 AND $\hat{\phi}$

For a sample of size N, the Karl Pearson formula for an $I \times J$ contingency table is given by

$$X_N^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{[X_{ij} - \hat{E}(X_{ij})]^2}{\hat{E}(X_{ij})}$$

Suppose that the sample is increased by a factor of C, so that $N_C = CN$. If this were done, it would be found that all the observed frequencies would be increased on the average by a factor of C. In addition, the estimated expected frequencies would also be increased on the average by a factor of C For the sample of size N_C, the Karl Pearson statistic would be given, on the average, by

$$\begin{aligned} X_{N_C}^2 &= \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{[CX_{ij} - C\hat{E}(X_{ij})]^2}{C\hat{E}(X_{ij})} \\ &= C\left[\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{[X_{ij} - \hat{E}(X_{ij})]^2}{\hat{E}(X_{ij})}\right] \\ &= CX_N^2 \end{aligned}$$

meaning that the Karl Pearson statistic would also be increased on the average by C However, the $\hat{\phi}$ coefficient would not be affected, since

$$\hat{\phi}^2_{N_C} = \frac{X^2_{N_C}}{N_C} = \frac{CX^2_N}{NC} = \frac{X^2_N}{N} = \hat{\phi}^2_N$$

This means that the statistical findings of Table 17-1 could have been documented with a sample of about one-eighth the size used, since for a sample of size $N_C = \frac{1}{8}(212) = 26.5 = 27$, the resulting $X^2_{N_C}$ would be about equal to $\frac{1}{8}(33\ 07) = 4.13$, which would have been greater than 3.84 and which would have led to rejection of H_0 In addition, the $\hat{\phi}$ would have been approximately equal to .3949 since it is not influenced by sample size.

As a further reinforcement of the meaning of these results, suppose a study with $I = 3$, $J = 4$, and $N = 72$ produced a Karl Pearson statistic with $X^2 = 4$. The hypothesis of statistical independence would not be rejected in this case since $X^2_6(.95) = 12\ 6$. For these data, $\hat{\phi} = .167$, which is indicative of a weak correlation that most researchers would be willing to equate with 0. Suppose, now, that the sample had been increased by a factor of five so that $N_C = 5(72) = 360$. With this large a sample, the approximate value of X^2 would be $(5)(4) = 20$ and the hypothesis of statistical independence would be rejected, even though $\hat{\phi} =\ 167$ had not changed.

This finding is true of all statistical procedures based upon statistics whose sampling distribution reduces in variability as sample size increases. In essence, this means that any population characteristic can be shown to be statistically significant, provided that a large enough sample is taken. This is the great weakness of statistical hypothesis testing and is why point and confidence-interval estimations have been employed in this book for analysis-of-variance studies and tests of homogeneity For an $I \times J$ contingency table, it is recommended that X^2 be given the minimal attention it is due and that the final decisions concerning the importance of a finding be dependent upon a consideration of $\hat{\phi}$, for this statistical variable contains the basic information about the correlation between Y_1 and Y_2, the variables of the study

It is perhaps true that much of the technical literature of the behavioral sciences is confusing or contradictory because of the homage paid to statistical testing of hypotheses and the resulting values of the test statistics. This is unfortunate, since many statistically significant differences reflect only the use of large samples and abnormally high statistical power While maximum power is a legitimate requirement of statistical testing, the practical and true meanings of statistical findings must not be ignored. As yet there are no statistical procedures that state when a finding is important. This decision is always made by the researcher. It is suggested that the decision of importance be based not on the numerical value of X^2 but upon $\hat{\phi}^2$ if the test is one of independence, or upon interesting confidence intervals if the test is one of homogeneity. In the same spirit, it follows that in the analysis of variance, the observed value of F is meaningless. What has greater meaning is the set of confidence intervals for the interesting contrasts of the parameters and the value

of $\hat{\omega}^2$ which measures the strength of the association between the criterion and treatment variables of the design

17-8 SUMMARY OF THE STATISTICAL TESTS AND CORRELATION MEASURES FOR CHI-SQUARE TESTS OF INDEPENDENCE

The basic tests of this chapter are summarized in Table 17-7. Since easy-to-use confidence intervals are not available for $\hat{\phi}$, no attempt is made to relate the test of hypothesis to the corresponding confidence-interval procedure.

17-9 SUMMARY

In this chapter, chi-square tests of independence were introduced and compared to chi-square tests of homogeneity. The simplest discriminating element between these two tests is that the test of independence is a one-sample problem while the test of homogeneity is a K-sample problem. Another distinguishing feature of the test of independence is that it is a bivariate rather than a univariate problem.

For an $I \times J$ contingency table in which Y_1 is defined by the set of mutually exclusive and exhaustive subsets $\{A_1, A_2, \ldots, A_I\}$ and in which Y_2 is defined by the corresponding set of mutually exclusive and exhaustive subsets $\{B_1, B_2, \ldots, B_J\}$, the test statistic for testing H_0: $P(A_i \cap B_j) = P(A_i)P(B_j)$ for all i and j is given by the classical Karl Pearson statistic:

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{[X_{ij} - \hat{E}(X_{ij})]^2}{\hat{E}(X_{ij})}$$

where X_{ij} is the observed frequency in the set $A_i \cap B_j$ and $\hat{E}(X_{ij}) = N\hat{P}(A_i)\hat{P}(B_j)$ is the estimated expected frequency of the set, given that H_0 is true. When H_0 is true, this statistic has a sampling distribution that can be approximated by chi-square with $\nu = (I - 1)(J - 1)$.

For the test of homogeneity, one has a single variable X that is defined by the set of mutually exclusive and exhaustive subsets $\{A_1, A_2, \ldots, A_I\}$, which is used to classify the elements of K independent random samples selected from K unrelated populations $P_1, P_2, \ldots, P_K$. The test statistic for testing H_0: $P(A_i | P_k) = P(A_i)$ for all i and k is given by the classical Karl Pearson statistic:

$$X^2 = \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{[X_{ki} - \hat{E}(X_{ki})]^2}{\hat{E}(X_{ki})}$$

where X_{ki} is the observed frequency of set A_i in the population P_k and where $\hat{E}(X_{ki}) = N_k\hat{P}(A_i)$ is the estimated expected frequency of the set, given that H_0 is true. When H_0 is true, this statistic has a sampling distribution that can be approximated by chi-square with $\nu = (K - 1)(I - 1)$.

Table 17-7. Statistical tests and correlation measures for chi-square tests of independence.

Case	*Hypothesis*	*Test statistic*	*Point estimate*	*Assumptions*
17	$H_0 . P(A \cap B) = P(A)P(B)$ or $\phi^2 = 0$ $H_1 : H_0$ is false	$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{[X_{ij} - \hat{E}(X_{ij})]^2}{\hat{E}(X_{ij})}$	$\hat{\phi} = \sqrt{\frac{X^2}{N}}$	1. Independence between pairs 2 Y_1 and Y_2 are binomial 3 Expected values greater than 5 4 One sample
18	$H_0 \cdot P(A_i \cap B_j) = P(A_i)P(B_j)$ or $\phi^2 = 0$ $H_1 . H_0$ is false $i = 1, 2, \ldots, I$ $j = 1, 2, \ldots, J$	$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{[X_{ij} - \hat{E}(X_{ij})]^2}{\hat{E}(X_{ij})}$	$\hat{\phi} = \sqrt{\frac{X^2}{NM}}$ where $M = \min I - 1, J - 1$	1 Independence between pairs 2. Y_1 and Y_2 are multinomial 3. Expected values greater than 5 4. One sample

For the chi-square test of homogeneity, one uses contrasts to locate sources of variance if the hypothesis of equal conditional probabilities is rejected, but for the chi-square test of independence, one estimates the strength of the association by means of the sample phi coefficient if the hypothesis of statistical independence is rejected. The simplest way to estimate the strength of the association is to compute

$$\hat{\phi} = \sqrt{\frac{X^2}{NM}}$$

where X^2 is the computed Karl Pearson statistic, N is the sample size, and M is the minimum of $I-1$ or $J-1$. When H_0 is rejected, it follows that $\phi \neq 0$, so that the test of H_0: $\phi = 0$ is identical to the test H_0: $P(A_i \cap B_j) = P(A_i)P(B_j)$.

The use of ϕ as a measure of the strength of association between two qualitative variables Y_1 and Y_2 was motivated by the notion of the covariance between two variables. For two discrete variables Y_1 and Y_2 such that Y_1 can assume the set of values $\{y_{11}, y_{12}, \ldots, y_{1I}\}$ and Y_2 can assume the set of values $\{y_{21}, y_{22}, \ldots, y_{2J}\}$ with $P[Y_1 = y_{1i} \cap Y_2 = y_{2j}]$, the covariance between the two variables is given by

$$\text{Cov}(Y_1, Y_2) = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{1i} - \mu_1)(y_{2i} - \mu_2) P[Y_1 = y_{1i} \cap Y_2 = y_{2j}]$$

So as to place this measure on a scale between -1 and $+1$ in the universe, the correlation coefficient is defined to be the following algebraic quantity

$$\phi = \frac{\text{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2}$$

While it is true that ϕ and $\text{Cov}(Y_1, Y_2)$ are measuring the same characteristic, it is customary to measure association via ϕ. This coefficient assumes the extreme values

Table 17-8. An example of near-perfect statistical association.

	A_1	A_2	A_3	A_4	*Total*
B_1	15	10	0	0	25
B_2	0	0	15	20	35
Total	15	10	15	20	60

of ± 1 when Y_1 and Y_2 are perfectly associated. Examples of perfect or near-perfect associations are shown in Tables 17-8 and 17-9. As can be seen, the distinguishing mark of a perfect association is that all the frequency of any row or column is localized in one cell of the row or column so that all remaining cells have zero frequency. A perfect correlation can be attained only if the contingency table is a square. For nonsquare contingency tables, only near-perfect correlations can be determined.

Table 17-9. Examples of perfect statistical association.

	A_1	A_2	A_3	*Total*
B_1	10	0	0	10
B_2	0	20	0	20
B_3	0	0	10	10
Total	10	20	10	40

	A_1	A_2	A_3	*Total*
B_1	0	10	0	10
B_2	10	0	0	10
B_3	0	0	20	20
Total	10	10	20	40

For a 2×2 contingency table such as Table 3-4, the sample phi coefficient is a simple function of N, $n(A)$, $n(B)$, and $n(A \cap B)$ and is given by

$$\hat{\phi} = \frac{Nn(A \cap B) - n(A)\,n(B)}{\sqrt{Nn(A) - [n(A)]^2}\,\sqrt{Nn(B) - [n(B)]^2}}$$

and in terms of the notation of Table 17-1,

$$\hat{\phi} = \frac{NX_{11} - X_{1\ } X_{\ 1}}{\sqrt{NX_{1\ } - X_{1\ }^2}\,\sqrt{NX_{\ 1} - X_{\ 1}^2}}$$

Finally, it was noted that while X^2 has the undesirable property of increasing with increases in N, $\hat{\phi}$ does not have that property and is therefore suggested as a statistic to employ in measuring meaningful differences or correlations. Since any population characteristic can be shown to be different or statistically significant from any specified numerical value by simple increases in sample size, it follows that unrestricted hypothesis testing is meaningless. For that reason, it is suggested that a researcher look at his data after performing a statistical test and decide whether a statistically significant finding is of practical or meaningful value. If the sample size is large, chances are that it is not. For example, if $X^2 = 10$, $N = 2{,}000$, and $\nu = 1$, then $\hat{\phi} = \sqrt{10/2000} = .05$ is not indicative of a meaningful statistical association even though the hypothesis H_0: $\phi = 0$ has been rejected. These statements are valid for all statistical tests and are worthy of thoughtful consideration and behavior.

Without doubt, a rational and universally accepted set of definitions of weak, moderate, and strong associations is not possible. The meaning attached to these adjectives varies with researchers and statisticians. For the individual new to these measures of association, the following guidelines are suggested:

Strength of association	*Range in* $\hat{\phi}$
Weak	$0 < \hat{\phi} < .33$
Moderate	$.33 < \hat{\phi} < .67$
Strong	$.67 < \hat{\phi} < 1.00$

In Chapter 18, a more complete discussion of this problem is presented.

EXERCISES

17-1. Suppose that two researchers performing independent studies with $\nu = 1$ reported the following statistics

	Study 1	*Study 2*
Value of X^2	15 3	189.6
Sample size	27	631

Which researcher has found a stronger relationship? What does this tell you about the effects of sample size in research investigations?

17-2. Compare and contrast chi-square tests of independence and homogeneity.

***17-3.** As part of a study on retired males over the age of sixty-five, 317 retired servicemen were asked how many years they served in the military and the number of times they were wounded in battle during their military career. Is the proper method of analysis of the resulting data a chi-square test of independence or a chi-square test of homogeneity? Defend your answer.

***17-4.** For the investigation of Exercise 17-3, is the method of contrasts an appropriate *post hoc* investigation procedure or is the determination of $\hat{\phi}$ a more reasonable *post hoc* method? Defend your answer.

***17-5.** When more than 20 percent of the estimated expected frequencies is less than 5, it is customary to pool the frequencies in neighboring cells. With this in mind, analyze the data of Exercise 3-10 in light of the methods presented in this chapter.

17-6. A physical education instructor at a large midwestern university measured the heights of 97 freshmen enrolled in a physical education class and asked each student if he had played on his high school basketball team. The results of the survey were as follows:

Response	*Height* 5′ 4″ TO 5′ 8″	5′ 9″ TO 6′	6′ 1″ TO 6′ 6″	*Total*
Yes	6	15	14	35
No	41	20	1	62
Total	47	35	15	97

Is this a test of independence or homogeneity? Why? What is the exact hypothesis of this investigation? Test this hypothesis. To what universe can this investigation be generalized?

***17-7.** In a study in which the effects of psychological alienation upon group participation in community activities was being investigated, 60 small business owners were given a test to measure their degree of alienation. Prior to the testing, each owner was ranked by two judges as to his activities in the community. The results were as shown

Degree of alienation	*Participation in community* LOW	MEDIUM	HIGH	*Total*
LOW	4	6	6	16
MEDIUM	6	10	5	21
HIGH	10	9	4	23
Total	20	25	15	60

On the basis of these data, can one conclude that participation is independent of alienation? What must you assume in making this test?

***17-8.** Could the study of Exercise 17-7 have been conducted with a smaller sample? For $\alpha = .05$, how large a sample could one have used to reach the same decision?

***17-9.** In the study of Exercise 17-7, 40 of the businesses had managers who were different from the owners. The statistics for these individuals are shown.

Degree of alienation	*Participation in community by owners*				*Participation in community by managers*			
	LOW	MEDIUM	HIGH	*Total*	LOW	MEDIUM	HIGH	*Total*
LOW	3	2	6	11	3	3	6	12
MEDIUM	5	7	2	14	10	5	3	18
HIGH	7	6	2	15	5	2	3	10
Total	15	15	10	40	18	10	12	40

In which group is the association between alienation and participation the strongest?

***17-10.** If there are no differences between the owners and the managers, the two sets of data in the table of Exercise 17-7 can be pooled to give the following statistics:

Degree of alienation	*Participation in the community*			
	LOW	MEDIUM	HIGH	*Total*
LOW	6	5	12	23
MEDIUM	15	12	5	32
HIGH	12	8	5	25
Total	33	25	22	80

Using the results of the pooling, estimate the expected frequencies for the cells of Exercise 17-9 and test the hypothesis that the distributions are identical. Are all assumptions for the use of chi-square satisfied? If not, how does one interpret the results? In any case, what is the correct number of degrees of freedom for this test?

18 INTRODUCTION TO STATISTICAL CORRELATION THEORY

Ours is said to be the age of anxiety, but what exactly is anxiety and how can it be measured? What are its manifestations and how does it affect the functioning of human beings? The initial difficulty in answering such questions—as in so many problems of psychology—is one of definition. . . we have been applying factor analysis to the problem of defining and learning how to measure anxiety.

. The factor analyst uses the correlation coefficient to calculate the precise extent to which two measured variables covary, or move about together. The coefficient can range from +1.0 through 0 to −1.0, indicating, respectively, a complete positive correlation, no relation at all and a complete negative, or inverse, correlation A correlation can be demonstrated visually by a "scatter plot." Each dot on the chart represents an observation of a patient at one session during which psychiatrists rated his anxiety level and a record was made of the degree to which his heart rate was varying. When these dissimilar measurements are reduced to "standard scores" and plotted on two coordinates, their pattern indicates the degree of correlation between "anxiety" and heart-rate variability. The fact that most of the dots fall into a rough ellipse (the closest approach to a straight line one can expect when many influences are at work) shows that there is a positive correlation between the two variables. The computation of a correlation coefficient from the actual anxiety ratings and heart-rate measurements gives a value of +.49, which is a fairly marked relation as psychological variables go.

18-1 THE BIVARIATE NORMAL DISTRIBUTION

The examples of the previous chapter involved the simultaneous measurement of two qualitative variables on each of the individual members of a single random sample selected from some universe or population of interest In those examples, the question of immediate interest centered on determining whether the two variables being studied were statistically independent in the populations from which the samples were drawn. For qualitative variables it was seen that tests based upon the chi-square distribution could be employed to assess the degree of independence. If the variables are measured on a discrete or continuous scale, the qualitative test procedures of the previous chapter are not appropriate. Instead, other methods of analysis are required. These other methods will be described and developed in this chapter.

For univariate statistical inference procedures, the normal distribution plays a prominent part. As a consequence, it might be thought that a two-variate analog to the normal distribution should play a similar role for statistical inference procedures involving bivariate continuous variables The probability distribution that assumes this position is called the *bivariate normal distribution.*

In the univariate model, the expected value and variance are used as parameters to characterize and define the major properties possessed by variables that have a normal probability distribution. In the bivariate normal model, each of the individual variables has univariate normal distribution. As might be expected, the parameters of the individual variables are also used as parameters of their parent bivariate distributions. In addition to these four parameters, a fifth parameter is required to define uniquely any specified bivariate normal distribution. This fifth parameter is called the correlation coefficient. In the technical literature, it is generally denoted by the Greek letter ρ (rho). This new parameter measures the degree to which the two random variables of the distribution fluctuate or vary together. In this sense, ρ is the continuous variable analog to ϕ. As might be expected, ρ and ϕ are defined under the same general framework. If the random variables of the bivariate normal distribution are denoted by Y_1 and Y_2, then any specific bivariate normal distribution will be denoted in this book by $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$. Thus, for example, $N(100, 50, 256, 100, \frac{1}{2})$ refers to the bivariate normal distribution with $E(Y_1) = \mu_1 = 100$, $E(Y_2) = \mu_2 = 50$, $\mathrm{Var}(Y_1) = \sigma_1^2 = 16^2$, $\mathrm{Var}(Y_2) = \sigma_2^2 = 10^2$, and $\rho = \frac{1}{2}$. In this example, it is clear that Y_1 refers to scores on the Stanford-Binet test of intelligence and Y_2 refers to a standardized variable with mean of 50 and standard deviation of 10.

18-2 THE GEOMETRY OF THE BIVARIATE NORMAL DISTRIBUTION

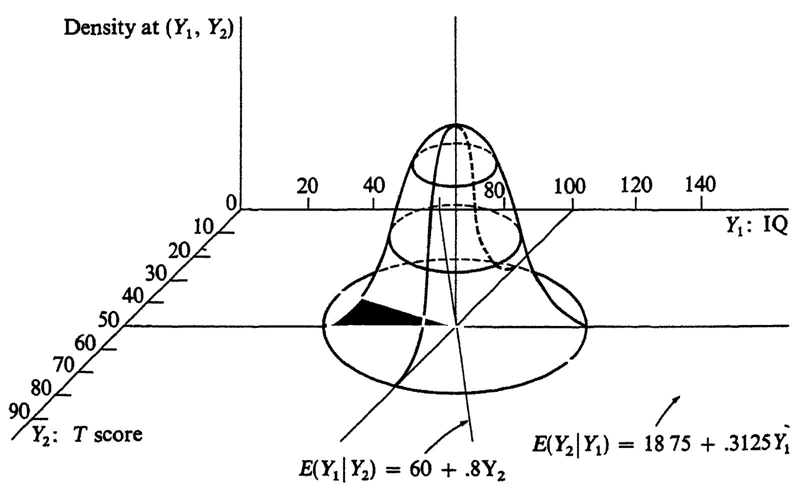
While a univariate normal distribution can be graphically represented by a two-dimensional drawing, a bivariate normal distribution requires a three-dimensional drawing. This three-dimensional nature is illustrated in Figure 18-1 for the bivariate normal distribution $N(100, 50, 256, 100, \frac{1}{2})$. The two perpendicular axes or reference system in the horizontal plane of the figure correspond to the numerical scale of values that each of the variables may assume, with one variable denoted by Y_1

and the other variable denoted by Y_2. As this drawing suggests, a bivariate normal distribution has the form of a distorted bell with major and minor axes at right angles to one another. From the diagram it appears that if planes parallel to the (Y_1, Y_2) plane but at different heights above the reference plane were to be drawn cutting the bell-shaped surface, ellipses would be generated with identical major axes parallel to a southwest-northeast direction, where north is identified with large numerical values of Y_2 and east is identified with large numerical values of Y_1. If the algebraic sign attached to ρ is negative, then the major axes of these ellipses lie in a northwest-southeast direction. Whenever $\rho = 0$, the resulting ellipses have major and minor axes that are exactly parallel to the two axes of reference As will be seen, if Y_1 and Y_2 are bivariate normal, $\rho = 0$ also means that Y_1 and Y_2 are statistically independent. If, in addition, $\sigma_1 = \sigma_2$, the ellipses degenerate into perfect circles.

In the univariate model, probabilities correspond to areas over intervals. As Figure 18-1 suggests, probabilities in the bivariate model correspond to volumes over areas. Unfortunately, these probabilities are difficult to compute. On the other hand, most research questions in the behavioral sciences that are based upon the bivariate normal model do not require the computations for such probabilities and for that reason these computations are not treated in this book.

In Figure 18-2 a vertical "slice" of the bivariate distribution at $Y_2 = 118$ is shown that appears to be normal in form. This slice is an example of one of the infinite number of conditional distributions that can be generated by simply passing a plane through the bell-shaped surface perpendicular to the (Y_1, Y_2) plane. All conditional distributions generated in this fashion are normal in form. When $Y_2 = 118$, the $E(Y_1|118) = 56.625$ and the $\mathrm{Var}(Y_1|118) = 192$. As would be expected,

Figure 18-1. The bivariate normal distribution $N(100, 50, 256, 100, \frac{1}{2})$



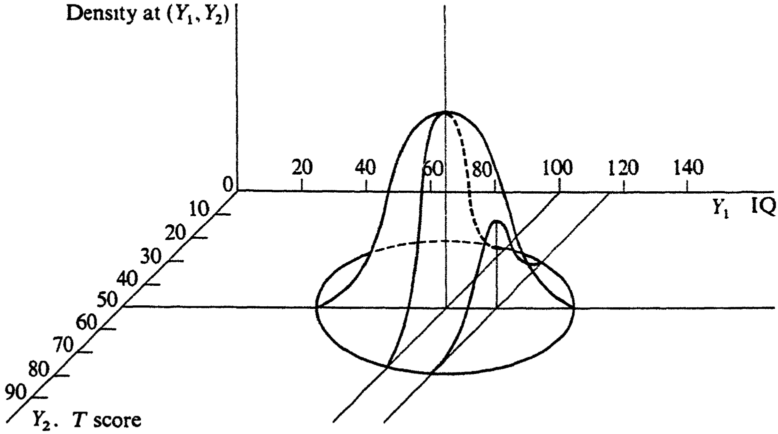


Figure 18-2. The bivariate normal distribution $N(100, 50, 256, 100, \frac{1}{2})$

these numerical values are related to the parameters of the parent bivariate normal distribution. The relationships that exist between the parameters of the parent bivariate distribution and the corresponding univariate distributions are summarized in the following theorem, which is stated without proof.

Theorem 18-1

The parameters of the conditional distributions defined for fixed values of Y_1 are given by

$$E(Y_2 \mid Y_1) = \mu_2 + \frac{\sigma_2}{\sigma_1}(Y_1 - \mu_1)$$

$$\mathrm{Var}(Y_2 \mid Y_1) = \sigma_2^2(1 - \rho^2)$$

and for fixed values of Y_2 are given by

$$E(Y_1 \mid Y_2) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(Y_2 - \mu_2)$$

$$\mathrm{Var}(Y_1 \mid Y_2) = \sigma_1^2(1 - \rho^2)$$

In essence, this theorem states that the conditional means of the bivariate normal distribution trace out as straight lines and that the variances of the conditional distributions are all equal to the same numerical value. This equality of variances is sometimes referred to as the property of *homoscedasticity*. For the bivariate normal distribution $N(100, 50, 256, 100, \frac{1}{2})$, the conditional variances are given by

$$\mathrm{Var}(Y_2 \mid Y_1) = 100[1 - (\tfrac{1}{2})^2] = 75$$

$$\mathrm{Var}(Y_1 \mid Y_2) = 256[1 - (\tfrac{1}{2})^2] = 192$$

As these results show, the variances remain constant for all values of Y_1 and Y_2. As Y_1 is allowed to vary from $-\infty$ to $+\infty$, all conditional distributions of Y_2 have their variance equal to 75, while if Y_2 is allowed to vary from $-\infty$ to $+\infty$, all conditional distributions of Y_1 have their variance equal to 192.

For this same distribution, the conditional expected values are given by

$$\begin{aligned} E(Y_2 \mid Y_1) &= 50 + (\tfrac{1}{2})(\tfrac{10}{16})(Y_1 - 100) \\ &= 50 + \tfrac{5}{16}(Y_1 - 100) \\ &= (50 - \tfrac{5}{16} \cdot 100) + \tfrac{5}{16} Y_1 \\ &= 18.75 + .3125\, Y_1 \end{aligned}$$

$$\begin{aligned} E(Y_1 \mid Y_2) &= 100 + (\tfrac{1}{2})(\tfrac{16}{10})(Y_2 - 50) \\ &= 100 + \tfrac{4}{5}(Y_2 - 50) \\ &= (100 - \tfrac{4}{5}\ 50) + \tfrac{4}{5} Y_2 \\ &= 60 + .8\, Y_2 \end{aligned}$$

As Y_1 is allowed to increase from $-\infty$ to $+\infty$, the conditional expectations of Y_2 also increase in a linear fashion. The same is true for the Y_2 variable. For example, as Y_1 increases by one unit, $E(Y_2 \mid Y_1)$ increases by 3125 unit, but as Y_2 increases by one unit, $E(Y_1 \mid Y_2)$ increases by .8 unit. These rates of change are illustrated in Table 18-1, which covers the ± 3 sigma range for both the Y_1 and the Y_2 variables.

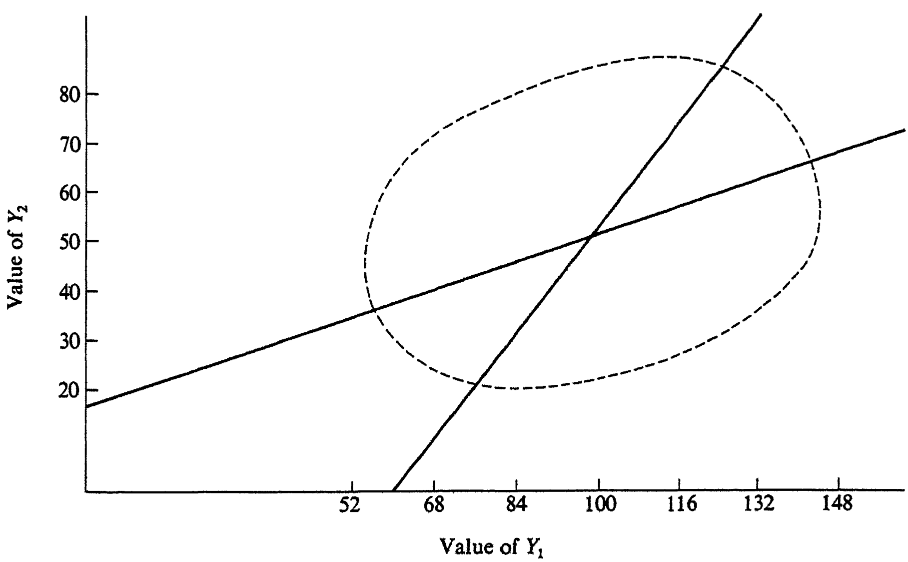
Table 18-1 Expected values and variances of the conditional distributions of $N(100, 50, 256, 100, \frac{1}{2})$ for sigma increases in each variable over a ± 3 standard deviation range.

Y_1	$E(Y_2 \mid Y_1)$	$\mathrm{Var}(Y_2 \mid Y_1)$	Y_2	$E(Y_1 \mid Y_2)$	$\mathrm{Var}(Y_1 \mid Y_2)$
52	35	75	20	76	192
68	40	75	30	84	192
84	45	75	40	92	192
100	50	75	50	100	192
116	55	75	60	108	192
132	60	75	70	116	192
148	65	75	80	124	192

While Table 18-1 gives a tabular presentation of the conditional expectations for certain isolated discrete values of Y_1 and Y_2, it is possible to show graphically all values of the conditional expectations for the major portion of the joint distribution by simply plotting the two equations $E(Y_2|Y_1) = 18.75 + .3125Y_1$ and $E(Y_1|Y_2) = 60 + .8Y_2$ in the (Y_1, Y_2) plane. These equations are graphed in Figure 18-3. A careful inspection will show that all pairs of Table 18-1 are on the plotted straight lines. It should be noted that the two lines intersect at $Y_1 = \mu_1 = 100$ and $Y_2 = \mu_2 = 50$. This point is called the *centroid* of the bivariate distribution and corresponds to the center of a univariate distribution. The two lines are called the *regression lines* of the bivariate distribution. The ellipse shows the 99.97 percent central region of the joint distribution.

The regression line of Y_2 on Y_1, the commonly employed name for the conditional expectations of the Y_2 variables, cuts the Y_2 axis at $Y_2 = 18.75$. This is called the Y_2 *intercept* and is generally denoted in the literature by $\alpha_{2.1}$. The Y_1 intercept for the regression line of Y_2 on Y_1 is given by $\alpha_{1.2} = 60$. For each unit increase in Y_1, the regression line of Y_2 on Y_1 increases by .3125. This constant increase in $E(Y_2|Y_1)$ for unit increase in Y_1 is called the *slope* of the regression line. In the literature it is generally denoted by $\beta_{2.1}$. For this distribution, $\beta_{2.1} = .3125$. The slope of the regression line of Y_1 on Y_2 is given by $\beta_{1.2} = .8$. As would be expected, the slopes and intercepts of the regression lines can be written in terms of the parameters of the parent bivariate distribution. The corresponding relationships are stated in the following theorems.

Figure 18-3. Regression lines for $N(100, 50, 256, 100, \frac{1}{2})$



Theorem 18-2

The intercepts and slopes of the regression lines are given by

$$\alpha_{2\ 1} = \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1$$

$$\alpha_{1\ 2} = \mu_1 - \rho\frac{\sigma_1}{\sigma_2}\mu_2$$

$$\beta_{2\ 1} = \rho\frac{\sigma_2}{\sigma_1}$$

$$\beta_{1\ 2} = \rho\frac{\sigma_1}{\sigma_2}$$

Proof. The intercept of the regression line of Y_2 on Y_1 is given by $E(Y_2|0)$. Thus

$$\alpha_{2.1} = E(Y_2|0) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(0 - \mu_1)$$

$$= \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1$$

In a like manner,

$$\alpha_{1\ 2} = \mu_1 - \rho\frac{\sigma_1}{\sigma_2}\mu_2$$

The slope of the regression line of Y_2 on Y_1 is given by

$$\beta_{2\ 1} = E(Y_2|\,Y_1 + 1) - E(Y_2|\,Y_1)$$

$$= \left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}\{(Y_1 + 1) - \mu_1\}\right] - \left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(Y_1 - \mu_1)\right]$$

$$= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}Y_1 + \rho\frac{\sigma_2}{\sigma_1} - \rho\frac{\sigma_2}{\sigma_1}\mu_1 - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}Y_1 + \rho\frac{\sigma_2}{\sigma_1}\mu_1$$

$$= \rho\frac{\sigma_2}{\sigma_1}$$

In a like manner,

$$\beta_{1.2} = \rho\,\frac{\sigma_1}{\sigma_2}$$

This completes the proof of the theorem.

As a corollary to the preceding theorem, it should be noted that

$$\beta_{1\,2}\beta_{2\,1} = \rho\frac{\sigma_1}{\sigma_2}\rho\frac{\sigma_2}{\sigma_1}$$
$$= \rho^2$$

The regression equations as defined in Theorem 18-1 are said to be in deviation form since $E(Y_2|Y_1) - \mu_2$ equals the deviation between the conditional mean and the unconditional mean, and $(Y_1 - \mu_1)$ measures the deviation between Y_1 and μ_1. Often the regression equations are written in what is termed the slope-intercept form. In this form the equations are as specified in Theorem 18-3.

Theorem 18-3

In slope-intercept form the regression lines can be written as

$$E(Y_2|Y_1) = \alpha_{2\,1} + \beta_{2\,1}Y_1$$
$$E(Y_1|Y_2) = \alpha_{1\,2} + \beta_{1\,2}Y_2$$

Proof.

$$E(Y_2|Y_1) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(Y_1 - \mu_1)$$
$$= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}Y_1 - \rho\frac{\sigma_2}{\sigma_1}\mu_1$$
$$= \left[\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1\right] + \rho\frac{\sigma_2}{\sigma_1}Y_1$$
$$= \alpha_{2\,1} + \beta_{2\,1}Y_1$$

In like manner,

$$E(Y_1|Y_2) = \alpha_{1\,2} + \beta_{1\,2}Y_2$$

This completes the proof. Incidentally, this proof also shows that

$$\alpha_{2\,1} = \mu_2 - \beta_{2\,1}\mu_1$$
$$\alpha_{1\,2} = \mu_1 - \beta_{1\,2}\mu_2$$

18-3 ESTIMATORS FOR THE PARAMETERS OF THE BIVARIATE NORMAL DISTRIBUTION

In general, the parameters of two variables known to be bivariate normal are unknown and must be estimated from a random sample selected from the parent universe. Let such a sample be denoted by the N pairs of numbers (y_{11}, y_{21}), (y_{12}, y_{22}),

$(y_{13}, y_{23}), \ldots, (y_{1N}, y_{2N})$ From univariate estimation theory it is known that unbiased estimates of μ_1, μ_2, σ_1^2, and σ_2^2 are given by

$$\bar{Y}_1 = \frac{1}{N}\sum_{i=1}^{N} y_{1i} \qquad S_1^2 = \sum_{i=1}^{N} \frac{(y_{1i} - \bar{Y}_1)^2}{N-1}$$

$$\bar{Y}_2 = \frac{1}{N}\sum_{i=1}^{N} y_{2i} \qquad S_2^2 = \sum_{i=1}^{N} \frac{(y_{2i} - \bar{Y}_2)^2}{N-1}$$

Thus, the only parameter left to estimate is ρ, the correlation coefficient While an estimator for ρ can be developed from advanced statistical principles, an intuitive procedure introduced in Chapter 17 for motivating the estimation of ϕ will be employed to reach the same formula. While this procedure is not mathematically rigorous, it should suffice for this book.

As stated in Section 17-4, the covariance of two discrete random variables is defined as

$$\text{Cov}(Y_1, Y_2) = \sum_{i=1}^{N} (y_{1i} - \mu_1)(y_{2i} - \mu_2) P[Y_1 = y_{1i} \cap Y_2 = y_{2i}]$$

If μ_1 and μ_2 are replaced by $\bar{Y}_1$ and $\bar{Y}_2$, and if $P[Y_1 = y_{1i} \cap Y_2 = y_{2i}]$ is replaced by $1/(N-1)$, then an unbiased estimate of $\text{Cov}(Y_1, Y_2)$ is given by

$$S_{12} = \frac{1}{N-1}\sum_{i=1}^{N} (y_{1i} - \bar{Y}_1)(y_{2i} - \bar{Y}_2)$$

Since it is customary to discuss correlation in terms that are free of the original measures, one normally divides the covariance by the product of the standard deviations. Following this procedure, we find that the correlation coefficient between Y_1 and Y_2 is given by

$$\rho = \frac{\text{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2}$$

Similar to the sample estimate of ϕ, a sample estimate of ρ is given by

$$\begin{aligned} r &= \frac{S_{12}}{S_1 S_2} \\ &= \frac{1}{S_1 S_2}\left[\frac{1}{N-1}\sum_{i=1}^{N} (y_{1i} - \bar{Y}_1)(y_{2i} - \bar{Y}_2)\right] \\ &= \frac{1}{N-1}\sum_{i=1}^{N} \left(\frac{y_{1i} - \bar{Y}_1}{S_1}\right)\left(\frac{y_{2i} - \bar{Y}_2}{S_2}\right) \end{aligned}$$

This estimate is called the sample correlation coefficient.

Once the parameters of the bivariate normal distribution have been determined, the parameters of the conditional distribution can also be estimated simply by substituting the estimated values in the parametric equations. The estimating equations are given by

$$\hat{Y}_1 = \bar{Y}_1 + r\frac{S_1}{S_2}(Y_2 - \bar{Y}_2)$$

$$\hat{Y}_2 = \bar{Y}_2 + r\frac{S_2}{S_1}(Y_1 - \bar{Y}_1)$$

and the conditional variances are given by

$$\hat{S}^2_{1\cdot 2} = S_1^2(1 - r^2)$$
$$\hat{S}^2_{2\cdot 1} = S_2^2(1 - r^2)$$

Since correlational studies are usually performed on large samples, the fact that $\hat{S}^2_{1.2}$ and $\hat{S}^2_{2\cdot 1}$ are biased is often ignored. Fortunately, unbiased estimators are available and can always be used for large or small samples. However, for correlational studies employing small sample sizes, the following unbiased estimators are to be preferred.

$$S^2_{2\cdot 1} = \frac{N-1}{N-2}S_2^2(1 - r^2)$$

and

$$S^2_{1.2} = \frac{N-1}{N-2}S_1^2(1 - r^2)$$

18-4 AN INTERPRETATION OF THE SAMPLE CORRELATION COEFFICIENT

If one lets

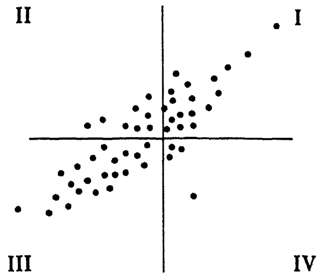
$$\hat{Z}_{1i} = \frac{y_{1i} - \bar{Y}_1}{S_1} \qquad \text{and} \qquad \hat{Z}_{2i} = \frac{y_{2i} - \bar{Y}_2}{S_2}$$

then the sample correlation coefficient reduces to

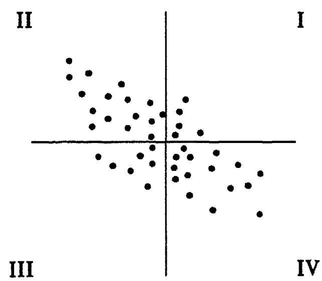
$$r = \frac{1}{N-1}\sum_{i=1}^{N}\hat{Z}_{1i}\hat{Z}_{2i}$$

This means that the sample correlation coefficient is an "average" value of the product of paired estimated Z values. To help interpret this measure, consider the three examples in Figure 18-4, which illustrate the scatter of three samples from universes in which Y_1 and Y_2 are said to be positively related, negatively related, and unrelated.

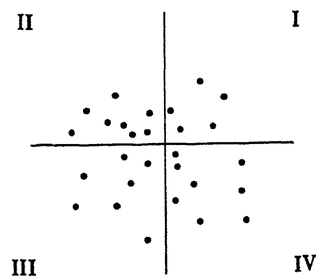
Each of the three graphs is called a *scatter diagram*. They represent the joint distribution of Y_1 and Y_2 of a random sample of size N selected from $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$. To aid the discussion, the average Y_1 and Y_2 values are used to partition the



Case I. Positively related variables



Case II. Negatively related variables



Case III Unrelated variables

Figure 18-4. Sample scatter diagrams for positively related, negatively related, and unrelated variables under the assumption that variables are bivariate normal.

scatter into four quadrants, which are numbered in a counterclockwise direction. With this partitioning, it is seen that if

1. $y_1 \in \{\text{I or IV}\}$, then $\hat{Z}_1$ is positive
2. $y_1 \in \{\text{II or III}\}$, then $\hat{Z}_1$ is negative
3. $y_2 \in \{\text{I or II}\}$, then $\hat{Z}_2$ is positive
4. $y_2 \in \{\text{III or IV}\}$, then $\hat{Z}_2$ is negative

Case I When Y_1 and Y_2 are positively related, the scatter of points extends along the major axis with a southwest-northeast direction with most of the points lying in quadrants I and III. In these quadrants, paired $\hat{Z}_1$ and $\hat{Z}_2$ have the same algebraic sign For example, in quadrant I, $\hat{Z}_1$ and $\hat{Z}_2$ are both positive, while in quadrant III they are both negative This means that in both quadrants, the products of paired $\hat{Z}_1$ and $\hat{Z}_2$ are always positive When the products are summed and divided by $N-1$, the resulting "average" product is positive, so that

$$r = \frac{1}{N-1} \sum_{i=1}^{N} \hat{Z}_{1i} \hat{Z}_{2i} > 0$$

For this reason, the relationship between Y_1 and Y_2 is said to be positive, as increases in Y_1 are accompanied by increases in Y_2.

Case II When Y_1 and Y_2 are negatively related, the scatter of points decreases as Y_1 increases. Thus, in the scatter diagram the points trace out a northwest-southeast pattern with most sample points contained in the II and IV quadrants. For these points, $\hat{Z}_1$ and $\hat{Z}_2$ have opposite algebraic signs so that their products are negative. Thus, the "average" value of the summed products will be negative, so that

$$r = \frac{1}{N-1} \sum_{i=1}^{N} \hat{Z}_{1i} \hat{Z}_{2i} < 0$$

Case III When Y_1 and Y_2 are completely unrelated, the scatter of points covers the four quadrants with equal probabilities Since each $\hat{Z}_{1i}\hat{Z}_{2i}$ product in quadrants I and II is counterbalanced by a corresponding product in quadrants III and IV, the average value of the products will be close to 0, and if there is no relationship whatsoever, then

$$r = \frac{1}{N-1} \sum_{i=1}^{N} \hat{Z}_{1i} \hat{Z}_{2i} = 0$$

It should be emphasized that this discussion has been based upon the assumption that Y_1 and Y_2 have a joint bivariate normal distribution. The reason for this assumption is illustrated by the scatter diagram shown in Figure 18-5. For this example, Y_1 and Y_2 are closely related. Knowledge of Y_1 entails considerable knowledge of Y_2, yet r for these data is very close to 0. This suggests that r is a measure of linear association that is insensitive to deviations from linearity. If the

relationship between Y_1 and Y_2 is not linear, then r is not an appropriate measure of the strength of the association.

It should be noted that curvilinear relationships are not uncommon in behavioral research. A correlation coefficient determined for such data will provide an estimate of the strength of the association that is too low. When the relationship is not linear, r should not be computed. Other measures are available and are discussed in more advanced textbooks. If a curvilinear relationship is encountered, then one should either consult an advanced statistics text or else seek the advice of a statistician.

The range of values of r is given by $-1 \leqslant r \leqslant +1$: $r=-1$ is associated with a perfect negative linear relationship $r=0$ is associated with no relationship; and $r=+1$ is associated with a perfect positive linear relationship. With most variables encountered in behavioral research, the typical range of r is given by $0 \leqslant r \leqslant +.8$. Correlation coefficients greater than .8 are quite unusual. Also, negative associations are not common.

When the association between Y_1 and Y_2 is a perfect linear association, every pair of observations lies on the same straight line and the value of $r=+1$. When $r=1$,

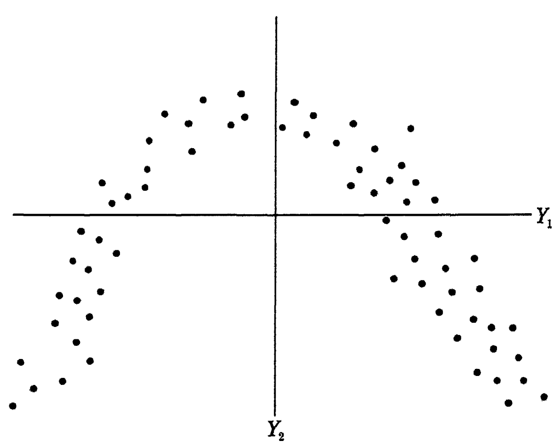
$$\hat{S}^2_{2.1} = S^2_1(1-r^2) = 0$$

which means that Y_2 has no variation for each value of Y_1. When $r=0$,

$$\hat{S}^2_{2.1} = S^2_1(1-r^2) = S^2_1$$

indicating that the variances of all conditional distributions are identical to the variance of the unconditional distribution. When there is a perfect negative associa-

Figure 18-5. Scatter diagram of two curvilinear related variables whose joint distribution is not bivariate normal.



tion between Y_1 and Y_2, then all points lie on the same line that decreases as Y_1 increases and the value of $r = -1$. Again,

$$\hat{S}^2_{2.1} = S^2_1(1 - r^2) = 0$$

showing that Y_2 has no variance for fixed Y_1.

18-5 EXPLAINED VARIANCE FOR CORRELATIONAL MODELS AS MEASURES OF RELATIONSHIP

One of the assumptions or properties of the bivariate normal correlation model is that the variances of all conditional distributions for fixed values of Y_1 and Y_2 are equal and are given by

$$\sigma^2_{2.1} = \sigma^2_2(1 - \rho^2)$$

and

$$\sigma^2_{1.2} = \sigma^2_1(1 - \rho^2)$$

In the discussion of this model, it was implicitly stated that the biased estimators

$$\hat{S}^2_{2.1} = S^2_2(1 - r^2)$$

and

$$\hat{S}^2_{1.2} = S^2_1(1 - r^2)$$

could be used as estimators for the corresponding parameters. For the discussion that follows, biased estimates will be used, not because they are more efficient but because their usage will simplify some of the equations and formulas to be derived. Furthermore, to avoid unnecessary repetitions, the discussion will be restricted to $\hat{S}^2_{2.1}$.

From the equation that defines $\hat{S}^2_{2.1}$ it follows that

$$\hat{S}^2_{2.1} = S^2_2 - r^2 S^2_2$$

so that

$$S^2_2 = \hat{S}^2_{2.1} + r^2 S^2_2$$

S^2_2 is a measure of variance that is based on the deviation of each individual outcome from the average of the group. In the terminology of analysis-of-variance designs, S^2_2 is a measure of the total variance in the observations. On the other hand, $\hat{S}^2_{2.1}$ is a measure of variance that is based on the deviation of each individual outcome from its best linear predicted value. In this sense it is a measure of residual or unexplained source of variance. Its similarity to the within measure of variance in analysis-of-variance designs is not surprising. This analogy with the analysis of variance further suggests that $r^2 S^2_2$ is a measure of variance related to the independent

variable of a correlation study, if there is an independent variable. Under this model, the last equation can be written as

(Total variance) = (unexplained variance) + (explained variance)

To give further meaning to this last equation, consider two outcomes of a correlation study for which the deviations from the group average are equal. Let the geometry of the model be as shown in Figure 18-6, and let the subscripts T, U, and E associated with the deviations be used with the end results in mind.

Consider the outcome (Y_{11}, Y_{21}) and its deviation from the group average. Let this deviation be partitioned into the two components as shown:

$$(Y_{21} - \bar{Y}_2) = (Y_{21} - \hat{Y}_{21}) + (\hat{Y}_{21} - \bar{Y}_2)$$

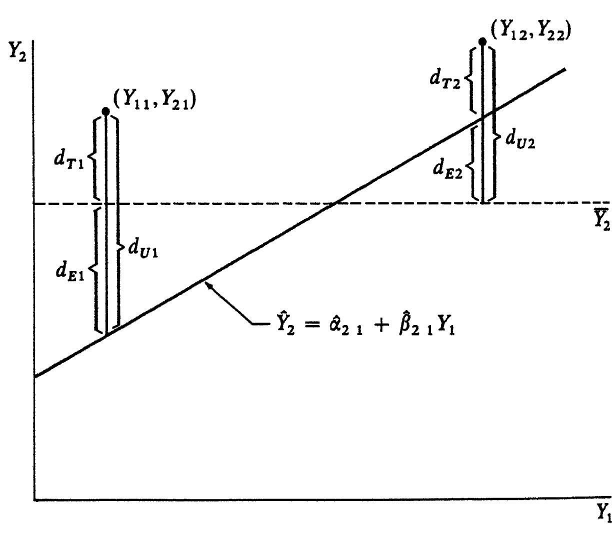
These deviations are easy to interpret in terms of the geometry of Figure 18-6:

$d_{T1} = Y_{21} - \bar{Y}_2$ = total deviation of the observed outcome from the group average (estimated deviation from the expected value if Y_1 and Y_2 were not correlated)

$d_{U1} = Y_{21} - \hat{Y}_{21}$ = deviation of the observed value from the best linear predicted value (estimated unexplained deviation from predicted value if Y_1 and Y_2 were correlated)

$d_{E1} = \hat{Y}_{21} - Y_2$ = deviation of the predicted value from the group average (estimated deviation of the predicted value of Y_2 from the average of the group)

Figure 18-6. Geometry of explained and unexplained variance for a bivariate correlation model.



An identical partitioning of the total deviation for the outcome (Y_{12}, Y_{22}) gives

$$(Y_{22} - \bar{Y}_2) = (Y_{22} - \hat{Y}_{22}) + (\hat{Y}_{22} - \bar{Y}_2)$$

For the illustrated outcomes, (Y_{12}, Y_{22}) is relatively closer to its predicted value than is (Y_{11}, Y_{21}). Since the total deviations from the group average are equal, it follows that

$$\frac{d_{U1}}{d_{T1}} > \frac{d_{U2}}{d_{T2}} \qquad \text{and} \qquad \frac{d_{E1}}{d_{T1}} < \frac{d_{E2}}{d_{T2}}$$

As these ratios suggest, most of the deviation from the group average for the first outcome relative to the total deviation must be attributable to chance or unexplained sources of variance and cannot be explained in terms of the correlation of Y_2 with Y_1 On the other hand, the deviation of the second outcome can easily be interpreted as correlational deviation or explained variance. As this suggests, if the correlation between Y_2 and Y_1 is relatively high, most of the outcomes will be close to the regression line, so that unexplained deviation d_{U1} will be small relative to their explained deviation d_{E1}, and even to their total deviation d_{T1}. On the other hand, if the correlation were weak, then one would expect the ratios to be large.

With these ideas in mind, consider the algebra of the total deviations. The similarity of this algebra to that of the analysis of variance is not accidental. By definition, the sum of squares total is given by

$$\text{SST} = \sum_{i=1}^{N} (y_{2i} - \bar{Y}_2)^2$$

Instead of subtracting and adding cell means and then expanding the resulting binomial, as one does in the analysis of variance, subtract and add both the predicted values. This gives

$$\text{SST} = \sum_{i=1}^{N} [(y_{2i} - \hat{Y}_{2i}) + (\hat{Y}_{2i} - \bar{Y}_2)]^2$$

With a little effort and relatively uninteresting algebra, it can be shown that

$$\sum_{i=1}^{N} (y_{2i} - \bar{Y})^2 = \sum_{i=1}^{N} (y_{2i} - \hat{Y}_{2i})^2 + \sum_{i=1}^{N} (\hat{Y}_{2i} - \bar{Y})^2$$

or that

$$\text{SST} = \text{SSU} + \text{SSE}$$

Consider the following ratio:

$$\hat{\omega}^2 = \frac{\text{SSE}}{\text{SST}} = \frac{\sum_{i=1}^{N} (\hat{Y}_{2i} - \bar{Y}_2)^2}{\sum_{i=1}^{N} (y_{2i} - \bar{Y}_2)^2} = \frac{\sum_{i=1}^{N} (\hat{Y}_{2i} - \bar{Y}_2)^2}{(N-1)\,S_2^2}$$

For the ith individual,

$$(\hat{Y}_{2i} - \bar{Y}_2) = [\bar{Y}_2 + \hat{\beta}_{2.1}(y_{1i} - \bar{Y}_1)] - \bar{Y}_2 = \hat{\beta}_{2.1}(y_{1i} - \bar{Y}_1)$$

so that

$$(\hat{Y}_{2i} - \bar{Y}_2)^2 = \hat{\beta}^2_{2.1}(y_{1i} - \bar{Y}_1)^2$$

and therefore

$$\sum_{i=1}^{N} (\hat{Y}_{2i} - \bar{Y}_2)^2 = \hat{\beta}^2_{2.1} \sum_{i=1}^{N} (y_{1i} - \bar{Y}_1)^2 = \hat{\beta}^2_{2.1}(N-1)\, S_1^2$$

Thus

$$\hat{\omega}^2 = \frac{\text{SSE}}{\text{SST}} = \frac{\hat{\beta}^2_{2.1}(N-1)\, S_1^2}{(N-1)\, S_2^2} = \hat{\beta}^2_{2.1} \frac{S_1^2}{S_2^2} = r^2 \frac{S_2^2\, S_1^2}{S_1^2\, S_2^2} = r^2$$

This last result states that the squared sample correlation coefficient is simply the ratio of the explained sum of squares to the total sum of squares.

To help interpret this remarkable result, consider the equation

$$S_2^2 = \hat{S}^2_{2.1} + r^2\, S_2^2$$

and the interpretation of this equation, which is given by

$$(\text{Total variance}) = (\text{unexplained variance}) + (\text{explained variance})$$

If the equation is divided by S_2^2, it follows that

$$\frac{S_2^2}{S_2^2} = \frac{\hat{S}^2_{2.1}}{S_2^2} + \frac{r^2\, S_2^2}{S_2^2}$$

or that

$$1 = (1 - r^2) + r^2$$

Since

$$(1 - r^2) = \frac{\hat{S}^2_{2.1}}{S_2^2} = \frac{(\text{unexplained variance})}{(\text{total variance})}$$

it makes intuitive sense to state that

$$r^2 = \frac{(\text{explained variance})}{(\text{total variance})}$$

In this form, r^2 can be interpreted as the proportion of the total variance in Y_2 that is explained by the correlation it shares with Y_1.

In a correlation study there are no dependent or independent variables. While this discussion has implicitly made this assumption, it should be noted that exactly the same results would have been obtained had the discussion been based upon $\hat{S}^2_{1.2}$ instead of $\hat{S}^2_{2.1}$.

In any case, it has been shown that r^2 is a measure of the amount of variance in one variable of a correlation model that can be attributed to the remaining variable. Furthermore, this discussion also suggests a simple way to interpret r via r^2. Without doubt, a correlation of .6 represents a stronger relationship than a correlation of .3. but how much stronger is not clear In terms of r^2, a correlation of .3 is known to explain 9 percent of the variability of one variable in terms of the remaining variable. For a correlation of .6, the amount of explained variance is 36 percent, four times that found for a correlation of .3. To further reinforce these notions, it is quite instructive to study the relationship between r, r^2, and $1 - r^2$. These relationships are summarized in Table 18-2 When $r = 0$, all conditional distributions are identical

Table 18-2. Relationship between r, r^2, and $1 - r^2$.

Value of r	*r^2 percent of explained variance*	*$1 - r^2$: percent of unexplained variance*	*Strength of the linear relationship*
0	.00	1.00	None
.1	01	.99	Very weak
.2	04	.96	
.3	09	.91	Weak
.4	16	.84	
.5	25	.75	Moderate
.6	.36	.64	Strong
7	49	.51	
.8	.64	.36	Very strong
.9	.81	.19	
1.0	1.00	.00	Perfect

with the same expected value and variance. This means that corresponding conditional probabilities are equal. Thus, as r deviates from 0, inequalities appear in the corresponding conditional probabilities.

From this table it is seen that when $r < .4$, more than 84 percent of the variance in one variable is still left unexplained. Even when $r = .7$, more than 50 percent of the total variance is still unexplained. In behavioral research, correlations exceeding .7 are rare. Typical correlations are found in the neighborhood of .4, which means that behavioral variables tend to have weak-to-moderate strength of associations. It should be stated that the subjective labeling of correlations as shown in Table 18-2 is not universally accepted by statisticians or behavioral researchers. The labeling is merely a convenience and should be used with care.

18-6 COMPUTATIONAL FORMULA FOR r

Just as there exist computational formulas for the sample variance, a convenient computational formula is available for r. This computational formula is stated in Theorem 18-4.

Theorem 18-4

The computational formula for r is given by

$$r = \frac{N\left(\sum_{i=1}^{N} y_{1i}y_{2i}\right) - \left(\sum_{i=1}^{N} y_{1i}\right)\left(\sum_{i=1}^{N} y_{2i}\right)}{\sqrt{N\left(\sum_{i=1}^{N} y_{1i}^2\right) - \left(\sum_{i=1}^{N} y_{1i}\right)^2}\sqrt{N\left(\sum_{i=1}^{N} y_{2i}^2\right) - \left(\sum_{i=1}^{N} y_{2i}\right)^2}}$$

Proof. By definition,

$$r = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{y_{1i}-\bar{Y}_1}{S_1}\right)\left(\frac{y_{2i}-\bar{Y}_2}{S_2}\right)$$

$$= \frac{1}{(N-1)S_1S_2}\sum_{i=1}^{N}(y_{1i}-\bar{Y}_1)(y_{2i}-\bar{Y}_2)$$

$$= \frac{1}{(N-1)S_1S_2}\sum_{i=1}^{N}(y_{1i}y_{2i} - y_{1i}\bar{Y}_2 - \bar{Y}_1y_{2i} + \bar{Y}_1\bar{Y}_2)$$

$$= \frac{1}{(N-1)S_1S_2}\left[\sum_{i=1}^{N} y_{1i}y_{2i} - \sum_{i=1}^{N} y_{1i}\bar{Y}_2 - \sum_{i=1}^{N}\bar{Y}_1y_{2i} + \sum_{i=1}^{N}\bar{Y}_1\bar{Y}_2\right]$$

$$= \frac{1}{(N-1)S_1S_2}\left[\sum_{i=1}^{N} y_{1i}y_{2i} - \bar{Y}_2\sum_{i=1}^{N} y_{1i} - \bar{Y}_1\sum_{i=1}^{N} y_{2i} + \bar{Y}_1\bar{Y}_2\sum_{i=1}^{N} 1\right]$$

$$= \frac{1}{(N-1)S_1S_2}\left[\sum_{i=1}^{N} y_{1i}y_{2i} - \bar{Y}_2(N\bar{Y}_1) - \bar{Y}_1(N\bar{Y}_2) + N\bar{Y}_1\bar{Y}_2\right]$$

$$= \frac{1}{(N-1)S_1S_2}\left[\sum_{i=1}^{N} y_{1i}y_{2i} - N\bar{Y}_1\bar{Y}_2\right]$$

$$= \frac{1}{(N-1)S_1S_2}\left[\sum_{i=1}^{N} y_{1i}y_{2i} - N\left(\frac{\sum_{i=1}^{N} y_{1i}}{N}\right)\left(\frac{\sum_{i=1}^{N} y_{2i}}{N}\right)\right]$$

$$= \frac{1}{(N-1)S_1S_2}\left[\frac{N\sum_{i=1}^{N} y_{1i}y_{2i} - \left(\sum_{i=1}^{N} y_{1i}\right)\left(\sum_{i=1}^{N} y_{2i}\right)}{N}\right]$$

Substituting the computing formulas for S_1 and S_2 into the right-hand side of the equation, we have

$$r = \frac{\dfrac{1}{N-1}\left[\dfrac{N\left(\sum_{i=1}^{N} y_{1i}y_{2i}\right) - \left(\sum_{i=1}^{N} y_{1i}\right)\left(\sum_{i=1}^{N} y_{2i}\right)}{N}\right]}{\sqrt{\dfrac{N\left(\sum_{i=1}^{N} y_{1i}^2\right) - \left(\sum_{i=1}^{N} y_{1i}\right)^2}{N(N-1)}}\sqrt{\dfrac{N\left(\sum_{i=1}^{N} y_{2i}^2\right) - \left(\sum_{i=1}^{N} y_{2i}\right)^2}{N(N-1)}}}$$

$$= \frac{N\left(\sum_{i=1}^{N} y_{1i}y_{2i}\right) - \left(\sum_{i=1}^{N} y_{1i}\right)\left(\sum_{i=1}^{N} y_{2i}\right)}{\sqrt{N\left(\sum_{i=1}^{N} y_{1i}^2\right) - \left(\sum_{i=1}^{N} y_{1i}\right)^2}\sqrt{N\left(\sum_{i=1}^{N} y_{2i}^2\right) - \left(\sum_{i=1}^{N} y_{2i}\right)^2}}$$

This completes the proof.

If the data have been grouped into a bivariate frequency table and if a code scale is introduced for the interval center, one should use the computational procedure illustrated in Exercise 18-4 at the end of this chapter.

18-7 AN EXAMPLE OF A CORRELATIONAL STUDY

As noted in Section 11-10, a common procedure used to increase the sensitivity of an experiment is to match subjects on relevant variables prior to instituting the experimental conditions. In a study on reading, 22 boys were matched with 22 girls on the basis of mental age and scores on the paragraph meaning section of the Stanford achievement test. Following eight months of special training, the subjects were tested for improved reading skills. Their improvement scores were as shown in Table 18-3. If the matching had been effective, then one would expect the scores to be associated in a linear fashion and their corresponding difference scores to be close to 0 in numerical value.

For the data of Table 18-3,

$$\sum_{i=1}^{22} y_{Bi} = 592 \qquad \sum_{i=1}^{22} y_{Bi}^2 = 16{,}582$$

$$\sum_{i=1}^{22} y_{Gi} = 655 \qquad \sum_{i=1}^{22} y_{Gi}^2 = 20{,}103$$

$$N = 22 \qquad \sum_{i=1}^{22} y_{Bi}\,y_{Gi} = 18{,}102$$

and

$$r = \frac{22(18102) - (592)(655)}{\sqrt{22(16582) - (592)^2}\sqrt{22(20103) - (655)^2}}$$

$$= .7608$$

Table 18-3. Scores on a reading test for 22 pairs of boys and girls matched on the basis of mental age and scores on the paragraph meaning section of the Stanford achievement test.

Pair	*Boy*	*Girl*	d_i	d_i^2
1	18	20	−2	4
2	17	26	−9	81
3	19	17	+2	4
4	26	30	−4	16
5	25	24	+1	1
6	23	31	−8	64
7	20	22	−2	4
8	19	30	−11	121
9	28	27	+1	1
10	33	32	+1	1
11	27	26	+1	1
12	28	29	−1	1
13	35	32	+3	9
14	31	35	−4	16
15	35	35	0	0
16	27	32	−5	25
17	29	35	−6	36
18	30	35	−5	25
19	29	35	−6	36
20	27	32	−5	25
21	34	35	−1	1
22	32	35	−3	9
Total	592	655	−63	481

Thus, it appears that the relationship between the scores of the boys and girls in a pair is quite strong and that the matching has been effective. In fact, most researchers would consider this to represent a strong positive correlation. For this correlation, $r^2 = .5796$ indicates that a strong linear association exists between the reading scores for the pairs matched on the basis of intelligence and school achievement measures.

The estimates of the remaining four parameters of the distribution are

$$\bar{Y}_B = \frac{\sum_{i=1}^{22} y_{Bi}}{N} = \frac{592}{22} = 26.9090$$

$$S_B^2 = \frac{N \sum_{i=1}^{22} y_{Bi}^2 - \left(\sum_{i=1}^{22} y_{Bi}\right)^2}{N(N-1)}$$

$$= \frac{22(16582) - (592)^2}{22(21)}$$

$$= 31.0389$$

$$S_B = 5.57$$

$$\bar{Y}_G = \frac{\sum_{i=1}^{22} y_{Gi}}{N} = \frac{655}{22} = 29.7727$$

$$S_G^2 = \frac{N \sum_{i=1}^{22} y_{Gi}^2 - \left(\sum_{i=1}^{22} y_{Gi}\right)^2}{N(N-1)}$$

$$= \frac{22(20103) - (655)^2}{22(21)}$$

$$= 28.6602$$

$$S_G = 5.35$$

The estimates of the conditional expected values are given by

$$\hat{Y}_B = \bar{Y}_B + r\frac{S_B}{S_G}(Y_G - \bar{Y}_G)$$

$$= 26.91 + .7608\ \tfrac{5.57}{5.35}\,(Y_G - 29.77)$$

$$= 3.33 + .7921(Y_G)$$

and

$$\hat{Y}_G = \bar{Y}_G + r\frac{S_G}{S_B}(Y_B - \bar{Y}_B)$$

$$= 29.77 + .7608\ \tfrac{5.35}{5.57}\,(Y_B - 26.91)$$

$$= 10.10 + .7308(Y_B)$$

The unbiased estimates of the conditional variances are given by

$$S_{G\ B}^2 = \frac{N-1}{N-2} S_G^2(1-r^2) = \tfrac{21}{20}(28.6602)(1-.7608^2) = 12.6753$$

$$S_{B\ G}^2 = \frac{N-1}{N-2} S_B^2(1-r^2) = \tfrac{21}{20}(31.0389)(1-.7608^2) = 13.7273$$

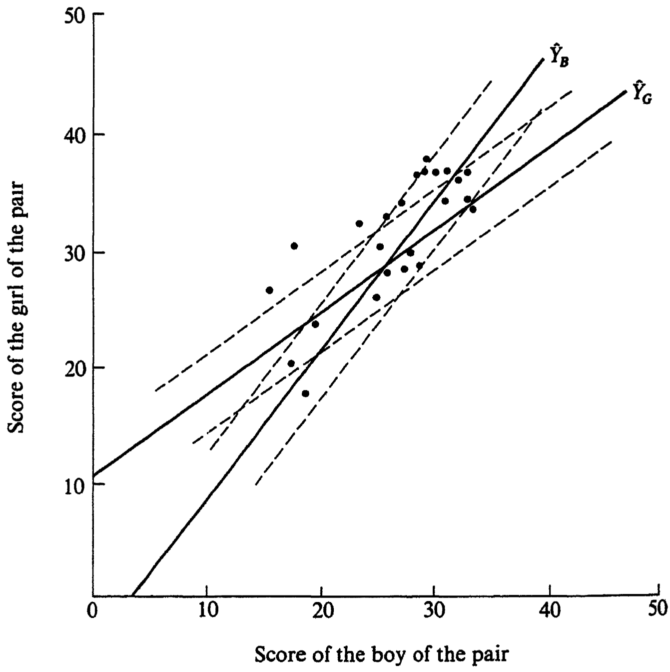
with

$$S_{G\ B} = 3.56 \qquad \text{and} \qquad S_{B\ G} = 3\ 70$$

The two regression lines with the associated one-standard deviation band are plotted in Figure 18-7. Also shown are the 22 observed pairs of values.

For the Y_G variables of the example, 15, or 68 percent of the cases, are in the ± 1 standard-deviation band. For a bivariate normal distribution, this is what one would expect For the Y_B variable, 16, or 72 percent of the cases, are in the ± 1 standard-deviation band. This too is within expectation. Furthermore, since in both directions the scatter diagram looks linear, it seems reasonable to assume that Y_B and Y_G have a joint bivariate normal distribution.

Figure 18-7. Scatter diagram for the bivariate set of 22 scores



18-8 THE MEANING OF A REGRESSION LINE

The dictionary meaning of regression suggests that regression is a retreat to a less advanced position. If such is the case, then it seems that a regression line should predict that the criterion measure or the dependent variable is at a less advanced level than is the predicting or independent measure. In a sense this is what the regression equation does.

To illustrate this, consider two extremely short individuals who marry and have children. While one would expect their children at completed growth to be short, one also expects them to be taller than either of the parents. On the other hand, consider two exceptionally tall individuals who marry and have children. While one would expect their children at completed growth to be tall, one also expects them to be shorter than either of the parents. It is said that the children's heights are regressed to the average height of the population. Thus, children of tall parents are closer to the average height than are their parents and children of short parents are closer to the average height than are their parents.

This phenomenon can be illustrated for the set of scores considered in Section 18-7. Consider the girl-boy pair in which the boy scored one standard deviation above the mean for the boys. The predicted score for his girl partner is given by

$$\hat{Y}_G = 10.10 + .7308(26.91 + 5.57) = 33.81$$

which in standard deviation units is equal to

$$\hat{Z}_G = \frac{33\ 81 - 29.77}{5.35} = .76$$

the value of the correlation coefficient. In a similar fashion, if a male score is one standard deviation below the average male score, then the predicted score for his female partner will be only .76 standard deviations below the average score for the girls. At the same time, a regression effect can be noted in the other direction. For example, a girl who scores two standard deviations above the mean expects her male counterpart to score $2r = 2(.76) = 1.52$ standard deviations above the mean of his group, while the girl who scores 1.5 standard deviations below the mean of her group expects her male counterpart to score $1.5r = 1.5(.76) = 1.14$ standard deviations below the mean for his group. In all these cases the predicted value is closer to the average of the mean of the paired individual than is the stated value from its own mean.

This regression to the mean is unavoidable in any experiment in which bivariate data are correlated. In fact, this procedure works in some very strange ways in behavioral research. An example where it entails serious methodological problems is the following, taken from a doctoral dissertation proposal that fortunately was corrected before the collection of data.

It was believed that the performance of Negro students in a literature program could be significantly improved by having them participate in the program. To show that the method was good, a controlled matched-pair study was anticipated. Both

white and Negro students were to be given an achievement test and then the subjects were to be matched on the basis of the test scores. As was expected, the Negro students scored in the lower part of the distribution with the whites scoring in the upper part Thus, when the matching was finally accomplished, only the top performing Negro students were included in the sample. Their matched partners were the poorest performing white students. If the treatment were to have no effect, one could predict the outcome of the study even before it was done. The white students on second testing will regress to the average of the white students and thereby experience a spurious increase in average performance. At the same time, the Negro students on second testing will also regress to the average of the Negro and thereby experience a spurious decrease in average performance A t test will show that the Negroes do poorly when in reality the treatment may have no effect Fallacies such as this one are very common. Unfortunately, experienced researchers fall into the making of this error. Matching is a very useful procedure for increasing the efficiency of a study However, there are appropriate and inappropriate methods for matching. In any case, a researcher is advised to discuss his design with a statistician before attempting to collect data. Many errors in research design can be avoided, and the literature will be less confusing.

18-9 SUMMARY

In this chapter, bivariate correlation theory was introduced via the bivariate normal distribution This distribution is characterized by five parameters: the two expected values of the two individual variables Y_1 and Y_2, the two variances of their corresponding univariate distributions, and ρ, the correlation coefficient between the two variables. This new parameter is related to the standard deviations of the univariate variables by means of

$$\rho = \frac{\operatorname{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2} = \frac{E(Y_1 - \mu_1)(Y_2 - \mu_2)}{\sigma_1 \sigma_2}$$

$$= E\left[\left(\frac{Y_1 - \mu_1}{\sigma_1}\right)\left(\frac{Y_2 - \mu_2}{\sigma_2}\right)\right]$$

In a random sample of size N selected from the bivariate normal distribution $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, an estimate of ρ is given by

$$r = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{y_{1i} - \bar{Y}_1}{S_1}\right)\left(\frac{y_{2i} - \bar{Y}_2}{S_2}\right)$$

which is simply computed from

$$r = \frac{N\left(\sum_{i=1}^{N} y_{1i}y_{2i}\right) - \left(\sum_{i=1}^{N} y_{1i}\right)\left(\sum_{i=1}^{N} y_{2i}\right)}{\sqrt{N\left(\sum_{i=1}^{N} y_{1i}^2\right) - \left(\sum_{i=1}^{N} y_{1i}\right)^2}\sqrt{N\left(\sum_{i=1}^{N} y_{2i}^2\right) - \left(\sum_{i=1}^{N} y_{2i}\right)^2}}$$

The remaining parameters of the bivariate normal distribution are estimated by $\bar{Y}_1$, $\bar{Y}_2$, S_1^2, and S_2^2, the usual sample mean and variances of the individual variables.

One property of the bivariate normal distribution that is of major interest in behavioral investigations is that the expected values of the complete set of conditional distributions that are defined for each value of Y_1 and Y_2 trace out straight lines that are referred to as linear regression equations These equations, in deviation form, are given by

$$E(Y_1 | Y_2) = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(Y_2 - \mu_2)$$

and

$$E(Y_2 | Y_1) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(Y_1 - \mu_1)$$

which are estimated in the sample by

$$\hat{Y}_1 = \bar{Y}_1 + r\frac{S_1}{S_2}(Y_2 - \bar{Y}_2)$$

and

$$\hat{Y}_2 = \bar{Y}_2 + r\frac{S_2}{S_1}(Y_1 - \bar{Y}_1)$$

When the regressions are stated in slope-intercept form,

$$E(Y_1 | Y_2) = \alpha_{1\ 2} + \beta_{1\ 2}\, Y_2$$

and

$$E(Y_2 | Y_1) = \alpha_{2\ 1} + \beta_{2.1}\, Y_1$$

which are estimated in the sample by

$$\hat{Y}_1 = \hat{\alpha}_{1\ 2} + \hat{\beta}_{1.2}\, Y_2$$

and

$$\hat{Y}_2 = \hat{\alpha}_{2\ 1} + \hat{\beta}_{2.1}\, Y_1$$

where

$$\hat{\alpha}_{1\ 2} = \bar{Y}_1 - r\frac{S_1}{S_2}\bar{Y}_2$$

$$\hat{\alpha}_{2.1} = \bar{Y}_2 - r\frac{S_2}{S_1}\bar{Y}_1$$

$$\hat{\beta}_{1.2} = r\frac{S_1}{S_2}$$

$$\hat{\beta}_{2.1} = r\frac{S_2}{S_1}$$

Another property of the bivariate normal distribution is that the variances of the conditional distributions are equal and given by

$$\sigma^2_{1.2} = \sigma^2_1(1 - \rho^2)$$

and

$$\sigma^2_{2.1} = \sigma^2_2(1 - \rho^2)$$

which are estimated in the sample by

$$\hat{S}^2_{1.2} = S^2_1(1 - r^2) \qquad \text{or} \qquad S^2_{1.2} = \frac{N-1}{N-2} S^2_1(1 - r^2)$$

and

$$\hat{S}^2_{2.1} = S^2_2(1 - r^2) \qquad \text{or} \qquad S^2_{2.1} = \frac{N-1}{N-2} S^2_2(1 - r^2)$$

When the variables are statistically independent, $\rho = 0$. When they are perfectly related so that an increase in Y_1 accompanies an increase in Y_2, $\rho = +1$. When they are inversely related in a perfect fashion so that decreases in Y_1 are accompanied by increases in Y_2, or so that increases in Y_1 are associated with decreases in Y_2, $\rho = -1$. Values of ρ between -1 and $+1$ correspond to various degrees of statistical association, provided that the relationships between Y_1 and Y_2 can be represented by straight lines. If the relationships are not linear, then ρ is not an appropriate measure of the strength of the association, since it will understate the true strength of the association.

EXERCISES

18-1. If sand is poured over a bull's-eye by means of a funnel, a bivariate normal distribution is produced over the target with the bull's-eye as a center. What is the value of ρ for the resulting mound of sand? Why?

18-2. (*a*) A correlation model always involves two linear regression equations. Why?
(*b*) If two variables are correlated, it does not necessarily follow that changes in one variable cause variations in the other variable. Produce some examples from behavioral research in which it can be concluded that correlation implies causation. At the same time, give some examples in which correlation does not imply causation.

18-3. It has been repeatedly shown that as the amount of smoking increases the amount of lung cancer and heart disease also increases. It has also been repeatedly shown that Negro students do not perform as well as Caucasian students on reading and arithmetic tests. Do these findings mean that smoking causes lung cancer and heart disease or that Negroes are genetically inferior to Caucasians when it comes to academic performance? Defend your answers.

***18-4.** Assign a Likert scale to the data of Exercise 3-10 and estimate the strength of the association between the two sets of responses Is there any evidence that $\rho = 0$? To simplify the computations of r, set up a work sheet like the one shown

Work sheet for computing r for grouped data.

Y_2	Y_1								
	A_1	A_2	A_3	A_4	f_j	Y_2	$f_j Y_2$	$f_j Y_2^2$	$f_{ij} Y_2 Y_1$
B_1	5	8	12	10	35	1	35	35	97
B_2	6	10	8	14	38	2	76	152	212
B_3	4	2	8	2	16	3			
B_4	1	2	1	7	11	4			
f_i	16	22	29	33	100				
Y_1	1	2	3	4					
$f_i Y_1$	16	44							
$f_i Y_1^2$	16	88							
$f_{ij} Y_1 Y_2$	33	76							

Note $33 = (1)(1)(5) + (1)(2)(6) + (1)(3)(4) + (1)(4)(1)$
$= 1[(1)(5) + (2)(6) + (3)(4) + (4)(1)] = 1[33] = 33$
$212 = (1)(2)(6) + (2)(2)(10) + (3)(2)(8) + (4)(2)(14)$
$= 2[(1)(6) + (2)(10) + (3)(8) + (4)(14)] = 2[106] = 212$

18-5. What is the connection between explained variance and conditional probability?

***18-6.** Following the t test of a comparison of a control condition versus an experimental condition, one can compute a correlation coefficient called the point biserial correlation. This correlation is the simple correlation coefficient between X and Y, where X is scored as 0 or 1, depending upon whether $X \in$ {control condition} or $X \in$ {experimental condition} For this model,

$$\sum_{i=1}^{N} x_i = N_1(0) + N_2(1) = N_2$$

$$\sum_{i=1}^{N} y_i = \sum_{i=1}^{N_1} y_{1i} + \sum_{i=1}^{N_2} y_{2i} = N_1 \bar{Y}_1 + N_2 \bar{Y}_2$$

$$\sum_{i=1}^{N} x_i^2 = N_1(0^2) + N_2(1^2) = N_2$$

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} y_i^2$$

$$\sum_{i=1}^{N} x_i y_i = (0) \sum_{i=1}^{N_1} y_{1i} + (1) \sum_{i=1}^{N_2} y_{2i} = N_2 \bar{Y}_2$$

Show that the point biserial correlation coefficient is given by

$$r = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\text{SST}}} \sqrt{\frac{N_1 N_2}{N_1 + N_2}}$$

***18-7.** Use the point biserial coefficient of Exercise 18-6 to estimate the strength of association for the data of Table 11-1. Does the estimate r suggest that $\rho \neq 0$? Defend your answer.

***18-8.** (*a*) Determine the correlation coefficient for the data of Table 5-8.
(*b*) What does it tell you about the matching process?

***18-9.** (*a*) Estimate the two regression lines for the data of Table 5-8.
(*b*) For student pair 4, $Y_1 = 27$ and $Y_2 = 23$. Estimate $\hat{Y}_1$ and $\hat{Y}_2$ for the pair.
(*c*) How do the estimated values compare?

***18-10.** (*a*) Estimate $\sigma_{1.2}^2$ for the data of Table 5-8 by means of $S_{1.2}^2$ and $\hat{S}_{1.2}^2$.
(*b*) Which estimate do you prefer? Why?

19 INFERENCE PROCEDURES FOR CORRELATED VARIABLES

This table leads to the first important result in the assessment of school factors associated with achievement: School-to-school variations in achievement, from whatever source (community differences, variations in the average home background of the student body, or variations in school factors), are much smaller than individual variations within the school, at all grade levels, for all racial and ethnic groups. This means that most of the variation in achievement could not possibly be accounted for by school differences, since most of it lies within the school.

PERCENT OF TOTAL VARIANCE IN INDIVIDUAL VERBAL ACHIEVEMENT SCORES THAT LIES BETWEEN SCHOOLS
(Corrected for degrees of freedom)

	Grades				
	12	9	6	3	1
Mexican Americans	20.20	15.87	28.18	24.35	23.22
Puerto Rican	22 35	21.00	31.30	26.65	16 74
Indian Americans	30.97	24.44	30.29	37.92	19.29
Oriental Americans	5 07	5.64	22.47	16.25	9.54
Negro South	22.54	20.17	22.64	34.68	23.21
Negro North	10.92	12.67	13 89	19.47	10.63
White South	10.11	9.13	11.05	17 73	18.64
White North	7.84	8.69	10.32	11.42	11.07

The table presents only the results for verbal achievement ... This result indicates that despite the wide range of diversity of school facilities, curriculum, and teachers, and despite the wide diversity among student bodies in different schools, over 70 percent of the variation in achievement for each group is variation within the same student body. The school-to-school difference is even less than overall figures for all groups indicate, because the school-to-school differences are generally least for Negroes and whites, and it is these groups which are numerically greatest Consequently, only about 10 to 20 percent of the total variation in achievement for the groups that are numerically most important lies between different schools.

From *Equality of Educational Opportunity*, by James S. Coleman, U.S. Department of Health, Education and Welfare Publication 38001, 1966

19-1 THE VARIANCE OF CORRELATED VARIABLES

While a correlation between Y_B and Y_G is expected for the study of Section 18-7, it is quite apparent that the study has been designed to test the hypothesis that boys and girls do equally well under the new teaching method. As shown in Section 13-4, the appropriate test statistic for testing H_0: $\theta = \mu_B - \mu_G = 0$ versus H_1: H_0 is false is given by the match pair t test. For these data,

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i = \frac{1}{22}(-63) = -2.8636$$

and

$$S_d^2 = \frac{N\left(\sum_{i=1}^{N} d_i^2\right) - \left(\sum_{i=1}^{N} d_i\right)^2}{N(N-1)}$$

$$= \frac{22(481) - (-63)^2}{22(21)} = 14.3139$$

so that

$$t = \frac{\sqrt{N}\bar{d}}{S_d} = \frac{\sqrt{22}(-2.8636)}{\sqrt{14.3139}} = -3.56$$

With $\alpha = .05$, H_0 should be rejected if $t < t_{21}(.025) = -2.080$ or if $t > t_{21}(.975) = +2.080$. Since $t = -3.56$, H_0 is rejected. If the students had not been matched, the appropriate test statistic, according to the theory of Section 13-2, is given by

$$t = \frac{\bar{Y}_B - \bar{Y}_G}{\sqrt{S_p^2/N_B + S_p^2/N_G}}$$

with $S_p^2 = \frac{1}{2}(31.0389 + 28.6602) = 29.8496$ and

$$t = \frac{-2.8636}{\sqrt{29.8496/22 + 29.8496/22}} = -1.74$$

In this case, H_0 would not be rejected. As this example shows, the effect of the high correlation upon the matching is to increase the power of the test and make it easier to reject H_0 when it is false. The reason for this increase in power is that when r is positive, $S_d^2 \leqslant S_p^2 + S_p^2 \leqslant S_B^2 + S_G^2$, as is shown in Theorem 19-1.

Theorem 19-1

When Y_1 and Y_2 are linearly related,

$$S_d^2 = \frac{1}{N-1}\sum_{i=1}^{N}(d_i - \bar{d})^2$$

$$= S_1^2 + S_2^2 - 2r\,S_1 S_2$$

Proof. For the ith pair, $d_i = y_{1i} - y_{2i}$, and

$$\begin{aligned}
\bar{d} &= \frac{1}{N}\sum_{i=1}^{N} d_i \\
&= \frac{1}{N}\sum_{i=1}^{N} (y_{1i} - y_{2i}) \\
&= \frac{1}{N}\left(\sum_{i=1}^{N} y_{1i} - \sum_{i=1}^{N} y_{2i}\right) \\
&= \frac{1}{N}(N\bar{Y}_1 - N\bar{Y}_2) \\
&= \bar{Y}_1 - \bar{Y}_2
\end{aligned}$$

By definition,

$$\begin{aligned}
S_d^2 &= \frac{\sum_{i=1}^{N}(d_i - \bar{d})^2}{N-1} \\
&= \frac{1}{N-1}\sum_{i=1}^{N} [(y_{1i} - y_{2i}) - (\bar{Y}_1 - \bar{Y}_2)]^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N} [(y_{1i} - \bar{Y}_1) - (y_{2i} - \bar{Y}_2)]^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N} [(y_{1i} - \bar{Y}_1)^2 - 2(y_{1i} - \bar{Y}_1)(y_{2i} - \bar{Y}_2) + (y_{2i} - \bar{Y}_2)^2] \\
&= \frac{\sum_{i=1}^{N}(y_{1i} - \bar{Y}_1)^2}{N-1} - \frac{2\sum_{i=1}^{N}(y_{1i} - \bar{Y}_1)(y_{2i} - \bar{Y}_2)}{N-1} + \frac{\sum_{i=1}^{N}(y_{2i} - \bar{Y}_2)^2}{N-1} \\
&= S_1^2 - 2rS_1S_2 + S_2^2
\end{aligned}$$

This completes the proof.

When S_d^2 is written in this form, it is easy to see that positively correlated variables are going to reduce the variance-of-difference variables. High positive correlations tend to reduce S_d^2 to an appreciable degree, while high negative correlations will inflate it.

Theorem 19-2

When $r > 0$, $S_d^2 < S_1^2 + S_2^2$.

Proof. By Theorem 19-1,

$$S_d^2 = S_1^2 + S_2^2 - 2rS_1S_2$$

When $r = 0$, $S_d^2 = S_1^2 + S_2^2$. Thus, for $r > 0$,

$$S_d^2 < S_1^2 + S_2^2$$

This completes the proof.

In like manner, when $r < 0$, $S_d^2 > S_1^2 + S_2^2$. When $r > 0$ and S_p^2 is substituted for S_1^2 and S_2^2, then $S_d^2 < S_p^2 + S_p^2 = 2S_p^2$. When $r < 0$ and S_p^2 is substituted for S_1^2 and S_2^2, then $S_d^2 > S_p^2 + S_p^2 = 2S_p^2$. These results are special cases of the following theorem, which is stated without proof.

Theorem 19-3

If $T = a_1 X_1 + a_2 X_2 + \cdots + a_K X_K$ and if the variables are not statistically independent, then

$$\begin{aligned}\text{Var}(T) = a_1^2 \text{Var}(X_1) + a_2^2 \text{Var}(X_2) + \cdots + a_K^2 \text{Var}(X_K) + 2[a_1 a_2 \text{Cov}(X_1, X_2)\\ + a_1 a_3 \text{Cov}(X_1, X_3) + \cdots + a_{K-1} a_K \text{Cov}(X_{K-1}, X_K)] \qquad \text{with } k_1 < k_2\end{aligned}$$

As can be seen, Theorem 6-4 is a special case of Theorem 19-3, which arises when all covariances are equal to 0 and each $a_i = 1$. On the other hand, Theorem 19-1 is the sample counterpart to Theorem 19-3 for $a_1 = +1$, $a_2 = -1$, and $a_3 = a_4 = \cdots = a_K = 0$. For these values of a, $\text{Var}(T) = \text{Var}(X_1 - X_2) = \text{Var}(X_1) + \text{Var}(X_2) - 2\text{Cov}(X_1, X_2) = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$.

As an example of the use of this theorem, consider the example of Section 7-10 in which two individuals are being compared concerning their performance on two different achievement tests For Mr. Brown of that example,

$$T = T_1 + T_2 = 38.75 + 40.83 = 79.58 = 79.5$$

while for Miss White,

$$T = T_1 + T_2 = 32.50 + 45.00 = 77.5$$

If the correlation between T_1 and T_2 is given by $\rho_{12} = .62$, the variance of $T = T_1 + T_2$ is given by

$$\begin{aligned}\text{Var}(T) &= \text{Var}(T_1) + \text{Var}(T_2) + 2\text{Cov}(T_1, T_2)\\ &= \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2\\ &= 100 + 100 + 2(62)(10)(10) = 324\end{aligned}$$

Since $E(T) = E(T_1) + E(T_2) = 50 + 50 = 100$, it follows that $T = 79.5$ corresponds to a percentile ranking of 13 because

$$\begin{aligned}P(T_1 \leqslant 79.5) &= P\left[Z \leqslant \frac{T - E(T)}{\sqrt{\text{Var}(T)}}\right] = P\left(Z \leqslant \frac{79.5 - 100}{18}\right)\\ &= P\left(Z \leqslant \frac{-20.5}{18}\right) = P(Z \leqslant -1.14) = .13\end{aligned}$$

On the other hand, $T = 77.5$ corresponds to a percentile ranking of 11, since

$$P(T_2 \leqslant 77.5) = P\left[Z \leqslant \frac{T - E(T)}{\sqrt{\operatorname{Var}(T)}}\right] = P\left(Z \leqslant \frac{77.5 - 100}{18}\right)$$

$$= P\left(Z \leqslant \frac{-22.5}{18}\right) = P(Z \leqslant -1.25) = .11$$

19-2 CONFIDENCE INTERVALS FOR THE PARAMETERS OF THE BIVARIATE NORMAL DISTRIBUTION

The confidence intervals for μ_1, μ_2, $\sigma_{1\ 2}^2$, and $\sigma_{2\ 1}^2$ are found in the usual fashion for univariate random variables except that the conditional standard deviations are used in determining the limits. In these cases, the confidence intervals are given by

$$\bar{Y}_1 - t_\nu\left(\frac{\alpha}{2}\right)\frac{S_{1.2}}{\sqrt{N}} < \mu_1 < \bar{Y}_1 + t_\nu\left(\frac{\alpha}{2}\right)\frac{S_{1.2}}{\sqrt{N}}$$

$$\bar{Y}_2 - t_\nu\left(\frac{\alpha}{2}\right)\frac{S_{2.1}}{\sqrt{N}} < \mu_2 < \bar{Y}_2 + t_\nu\left(\frac{\alpha}{2}\right)\frac{S_{2.1}}{\sqrt{N}}$$

$$\frac{(N-2)\,S_{1.2}^2}{\chi_\nu^2(1-\alpha/2)} < \sigma_{1\ 2}^2 < \frac{(N-2)\,S_{1\ 2}^2}{\chi_\nu^2(\alpha/2)}$$

$$\frac{(N-2)\,S_{2.1}^2}{\chi_\nu^2(1-\alpha/2)} < \sigma_{2\ 1}^2 < \frac{(N-2)\,S_{2\ 1}^2}{\chi_\nu^2(\alpha/2)}$$

with $\nu = N - 2$.

The confidence interval for ρ requires a further development of theory, since the sampling distribution of r is not symmetrical unless $\rho = 0$. The exact sampling distribution of r is quite complex, but fortunately there exists a transformation developed by Sir Ronald Fisher, commonly known as Fisher's z transformation, which converts the distribution to an approximate normal form. The transformation is given by

$$z = \frac{1}{2}\log_e\frac{1+r}{1-r}$$

The important properties of this variable are summarized in Theorem 19-4, which is stated without proof.

Theorem 19-4

The transformed Fisher variable has a distribution that is close to normal in form with expected value and variance given approximately by

$$E(z) = \frac{1}{2}\log_e\frac{1+\rho}{1-\rho} \qquad \operatorname{Var}(z) = \frac{1}{N-3}$$

An approximate $(1-\alpha)$ percent confidence interval for ρ can be determined from

$$z - Z\left(\frac{\alpha}{2}\right)\sigma_z < E(z) < z + Z\left(\frac{\alpha}{2}\right)\sigma_z$$

by transforming back to the original scale of r.

To simplify the determination of any confidence interval for ρ, the transformation of r to z is presented in Table A-12. The use of this table is illustrated for the data of Section 18-7. For these data, $r = .76$. From Table A-12 it is seen that when $r = .76$, the value of $z = .99621$. Thus, the 95 percent confidence interval for $E(z)$ is given by

$$.99621 - 1.96\frac{1}{\sqrt{22-3}} < E(z) < 99621 + 1.96\frac{1}{\sqrt{22-3}}$$

$$99621 - .4479 < E(z) < .99621 + .4479$$

$$54831 < E(z) < 1.44411$$

Reading "backward" in the table, the confidence interval for ρ is given by

$$.495 < \rho < .895$$

On the basis of this interval, there is good reason to believe that in the population $\rho > .50$, a fairly moderate to strong degree of correlation.

19-3 TESTS OF HYPOTHESIS FOR CORRELATION COEFFICIENTS

Since one-sample confidence intervals and tests of hypothesis are interchangeable, it makes sense to use Fisher's z to test hypotheses about ρ. To test $H_0\colon \rho = \rho_0$, the appropriate test statistic is given by

$$Z = \frac{z - E(z)}{\sigma_z} = \frac{\frac{1}{2}\log_e\frac{1+r}{1-r} - \frac{1}{2}\log_e\frac{1+\rho}{1-\rho}}{1/\sqrt{N-3}}$$

which is approximately $N(0,1)$. For this test, Table A-12 would again be used to compute z and $E(z)$.

To test $H_0\colon \rho_1 = \rho_2$ versus $H_1\colon H_0$ is false, where ρ_1 and ρ_2 are the parameters of two independent bivariate normal populations, the appropriate test statistic is given through the usual two-sample theory. If $\theta = E(z_1) - E(z_2)$, then $\hat{\theta} = z_1 - z_2$, and

$$\begin{aligned}\operatorname{Var}(\hat{\theta}) &= \operatorname{Var}(z_1) + \operatorname{Var}(z_2)\\ &= \frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}\end{aligned}$$

Thus, one can test the equality of population correlations by means of the following test statistic:

$$Z = \frac{z_1 - z_2}{\sqrt{1/(N_1 - 3) + 1/(N_2 - 3)}}$$

which is approximately $N(0,1)$ when H_0: $\rho_1 = \rho_2$ is true In a corresponding manner, the $(1 - \alpha)$ percent confidence interval for $E(z_1) - E(z_2)$ is given by

$$(z_1 - z_2) - Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}} < E(z_1) - E(z_2) < (z_1 - z_2) + Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

In this case, it is not customary to transform back to the individual ρ values because the arithmetic computations are not simple. As a result, if the above confidence interval does not include 0, then one can conclude that in the parent populations, $\rho_1 \neq \rho_2$.

As an example of the use of this result, consider a hypothetical study in which mental age is correlated with score on an achievement test, in the control group, $r_C = .32$, while in the experimental group, $r_E = .56$, and $N_C = N_E = 53$. For these data, $z_C = .33165$ and $z_E = .63283$, so that the value of the test statistic for testing H_0: $\rho_C = \rho_E$ versus H_1: $\rho_C \neq \rho_E$ is given by

$$Z = \frac{z_C - z_E}{\sqrt{1/(N_C - 3) + 1/(N_E - 3)}} = \frac{.33165 - 63283}{\sqrt{\frac{1}{50} + \frac{1}{50}}} = \frac{-30118}{.2} = -1.56$$

With $\alpha = .05$, the decision rule is to reject H_0 if $Z < -1.96$ or $Z > 1.96$. In this case H_0 is not rejected.

Since H_0 has not been rejected, one can obtain a more efficient estimate of ρ by working "backward" in the following equation:

$$Z = \frac{(N_C - 3)\, z_C + (N_E - 3)\, z_E}{(N_C - 3) + (N_E - 3)}$$

For these data,

$$Z = \frac{50(.33165) + 50(.63283)}{50 + 50} = .48224$$

so that the estimate of the common correlation coefficient is given by $r = .45$.

In behavioral research it is common to select K independent random samples from K independent bivariate normal distributions and then ask the question as to whether the correlation coefficients of the K different distributions are equal. A simple test of this hypothesis can be constructed via z and the chi-square distribution. This test is based upon the same theory used to develop F^*, the F test for universes with unequal variances.

As a first step in deriving this test, one must estimate $E(z)$ under the hypothesis that all correlation coefficients are equal. This estimate is given in Theorem 19-5.

Theorem 19-5

Under the hypothesis of equal ρ values, an estimate of $E(z)$ is given by

$$z_0 = \frac{\sum_{k=1}^{K} \frac{1}{\sigma_{z_k}^2} z_k}{\sum_{k=1}^{K} \frac{1}{\sigma_{z_k}^2}} = \frac{\sum_{k=1}^{K} (N_k - 3) z_k}{\sum_{k=1}^{K} (N_k - 3)}$$

Finally, the appropriate test statistic for testing equal correlations is stated in Theorem 19-6.

Theorem 19-6

The test statistic for the hypothesis of equal correlations is given by

$$U_0 = \sum_{k=1}^{K} \left(\frac{z_k - z_0}{\sigma_{z_k}} \right)^2 = \sum_{k=1}^{K} (N_k - 3)(z_k - z_0)^2$$

where U_0 is approximately chi-square with $\nu = K - 1$ degrees of freedom.

Proof. For the first sample,

$$Z_1 = \frac{z_1 - E(z_1)}{\sigma_{z_1}} \qquad \text{is } N(0,1)$$

For the second sample,

$$Z_2 = \frac{z_2 - E(z_2)}{\sigma_{z_2}} \qquad \text{is } N(0,1)$$

and for the Kth sample,

$$Z_K = \frac{z_K - E(z_K)}{\sigma_{z_K}} \qquad \text{is } N(0,1)$$

Thus, by Theorem 10-5,

$$U_1 = Z_1^2 + Z_2^2 + \cdots + Z_K^2 = \sum_{k=1}^{K} \left[\frac{z_k - E(z_k)}{\sigma_{z_k}} \right]^2$$

is approximately χ^2 with K degrees of freedom. If z_0 is substituted for $E(z_k)$, then 1 degree of freedom is employed to restrict the value of U_1, so that

$$U_0 = \sum_{k=1}^{K} (N_k - 3)(z_k - z_0)^2$$

is approximately χ^2 with $\nu = K - 1$ degrees of freedom. This completes the proof.

The use of this test and the corresponding *post hoc* procedure is illustrated by an example. This example, with $K = 5$, is based on correlating test scores between two standardized tests given to fifth-grade children in five urban elementary schools in the Oakland, California, Unified School District. The two tests are the Kuhlmann–Anderson intelligence and the paragraph meaning section of the Stanford achievement test. The sample sizes, the correlation coefficients and z values are shown in Table 19-1. As can be seen, the correlation coefficients differ from one

Table 19-1. Computations required for the determination of U_0 for testing H_0: $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5$.

Statistic	*School 1*	*School 2*	*School 3*	*School 4*	*School 5*
r_k	66	70	68	92	.44
N_k	58	68	113	37	91
z_k	.793	867	829	1 589	.472
$\text{Var}(z_k)$	.0182	0154	0091	0294	.0114
$\dfrac{1}{\text{Var}(z_k)}$	55	65	110	34	88

another. Since these particular schools were located in neighborhoods spreading over broad socioeconomic strata, this variance in correlations is not unexpected. Using the techniques presented here, we can make a test of this assumption. In particular, the hypothesis

$$H_0\colon \rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = \rho_0$$

is tested against the alternative hypothesis that H_0 is false.

Computations for the test are summarized in Table 19-1. For this example,

$$z_0 = \frac{55(.793) + 65(.867) + 110(.829) + 34(1.589) + 88(.472)}{55 + 65 + 110 + 34 + 88}$$

$$= .814$$

and

$$U_0 = 55(.793 - .814)^2 + 65(.867 - .814)^2 + 110(.829 - .814)^2 + 34(1.589 - .814)^2 + 88(.472 - .814)^2$$

$$= 30.94$$

Since five correlation coefficients are being compared, U_0 is approximately chi-square with 4 degrees of freedom if H_0 is true. For $\alpha = .05$, $\chi_4^2(.95) = 9.49$, and, as a result, H_0 is rejected. As a result, it follows that at least two of the correlations are different or that some contrast is significantly different from 0. In this case, the interesting and most meaningful contrasts are the simple differences between the

correlations. The general form of the $(1-\alpha)$ percent simultaneous confidence intervals about the simple contrasts is given by

$$\psi = (z_{k_1} - z_{k_2}) \pm \sqrt{\chi^2_{K-1}(1-\alpha)}\sqrt{\frac{1}{N_{k_1}-3} + \frac{1}{N_{k_2}-3}}$$

For this example,

$$S^* = \sqrt{\chi^2_{K-1}(1-\alpha)} = \sqrt{\chi^2_4(.95)} = \sqrt{9.49} = 3.08$$

The simple contrasts are as follows:

$$
\begin{array}{rcll}
-.64 < & E(z_1) - E(z_2) & < .49 & \text{not significant} \\
-.54 < & E(z_1) - E(z_3) & < .47 & \text{not significant} \\
-.47 < & E(z_1) - E(z_4) & < -.12 & \text{significant} \\
-.21 < & E(z_1) - E(z_5) & < .85 & \text{not significant} \\
-.44 < & E(z_2) - E(z_3) & < .52 & \text{not significant} \\
-1.37 < & E(z_2) - E(z_4) & < -.07 & \text{significant} \\
-.11 < & E(z_2) - E(z_5) & < .90 & \text{not significant} \\
-1.36 < & E(z_3) - E(z_4) & < -.16 & \text{significant} \\
-.08 < & E(z_3) - E(z_5) & < .80 & \text{not significant} \\
.50 < & E(z_4) - E(z_5) & < 1.74 & \text{significant}
\end{array}
$$

On the basis of these 10 confidence intervals, it is now concluded that School 4 differs from the others. *Post hoc* inspection of the sample statistics suggests that the correlations for Schools 1, 2, and 3 are equal and as a group might differ from the correlation for School 5. This *post hoc* hypothesis can be tested as follows: a contrast associated with this hypothesis is

$$
\begin{aligned}
\hat{\psi} &= \tfrac{1}{3}(z_1 + z_2 + z_3) - z_5 \\
&= \tfrac{1}{3}(.793 + .867 + .829) - .472 = .358
\end{aligned}
$$

and the $\text{Var}(\hat{\psi})$ is given by

$$
\begin{aligned}
\text{Var}(\hat{\psi}) &= \tfrac{1}{9}\text{Var}(z_1) + \tfrac{1}{9}\text{Var}(z_2) + \tfrac{1}{9}\text{Var}(z_3) + \text{Var}(z_5) \\
&= \tfrac{1}{9}(.0182) + \tfrac{1}{9}(.0154) + \tfrac{1}{9}(.0091) + .0114 = .0161
\end{aligned}
$$

The 95 percent confidence interval for ψ is

$$
\begin{gathered}
\hat{\psi} - \sqrt{\chi^2_4(.94)}\sqrt{\text{Var}(\hat{\psi})} < \psi < \hat{\psi} + \sqrt{\chi^2_4(.95)}\sqrt{\text{Var}(\hat{\psi})} \\
.358 - 3.08\sqrt{.0161} < \psi < .358 + 3.08\sqrt{.0161} \\
-.032 < \psi < .748
\end{gathered}
$$

Since this confidence interval includes 0, it is concluded that the *post hoc* hypothesis concerning these four schools is not supported.

In this example, 11 individual confidence intervals have been investigated with an over-all probability of a type I error controlled at .05. If each confidence interval had been determined with $Z = \pm 1.96$, the familiar normal curve values, the probability of at least one type I error among the 11 intervals would be less than or equal to $\alpha \leqslant 11(.05) = .55$, a rather high probability of error. When the S^* value of 3.08 is used, the intervals will be wide, but more important, the corresponding probability of at least one type I error is less than or equal to .05.

19-4 SPEARMAN'S RANK CORRELATION COEFFICIENT

Oftentimes correlation coefficients are used to measure the amount of agreement in two sets of measures. An example would be found in trying to determine whether the sports analysts of the Associated and United Press measure the same thing when they rank football teams. Since rankings are on a discrete scale with a uniform probability distribution, using normal bivariate distribution theory is somewhat questionable. To overcome this problem, one can use Spearman's rank correlation coefficient to estimate the degree of agreement in two lists of numbers. The Spearman rank correlation coefficient is derived in Theorem 19-7.

Theorem 19-7

The Spearman rank correlation coefficient is given by

$$r_S = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N^3 - N}$$

where

$$d_i = (R_{1i} - R_{2i})$$
$$R_{1i} = \text{rank of } y_{1i}$$
$$R_{2i} = \text{rank of } y_{2i}$$

Proof. In beginning algebra courses, it is usually shown that

$$\sum_{i=1}^{N} R_i = 1 + 2 + 3 + \cdots + N = \frac{N}{2}(N+1)$$

and that

$$\sum_{i=1}^{N} R_i^2 = 1^2 + 2^2 + 3^2 + \cdots + N^2 = \frac{N}{6}(N+1)(2N+1)$$

If one has a sample consisting of $(1,2,3, \ldots, N)$ its variance is given by

$$S_R^2 = \frac{N\left(\sum_{i=1}^{N} R_i^2\right) - \left(\sum_{i=1}^{N} R_i\right)^2}{N(N-1)}$$

$$= \frac{N[(N/6)(N+1)(2N+1)] - [(N/2)(N+1)]^2}{N(N-1)}$$

$$= \frac{N^2(N+1)}{N(N-1)}\left[\frac{2N+1}{6} - \frac{N+1}{4}\right]$$

$$= \frac{N(N+1)}{N-1}\left[\frac{8N+4-6N-6}{24}\right]$$

$$= \frac{N(N+1)}{24(N-1)}[2N-2]$$

$$= \frac{N(N+1)(N-1)}{12(N-1)}$$

$$= \frac{N(N+1)}{12}$$

If ranks are assigned to the observed Y_1 and Y_2, it then follows that $S_{R_1}^2 = S_{R_2}^2 = S_R^2$. By Theorem 19-1,

$$S_d^2 = S_{R_1}^2 + S_{R_2}^2 - 2r_S S_{R_1} S_{R_2}$$

$$= S_R^2 + S_R^2 - 2r_S S_R^2$$

$$= 2S_R^2(1 - r_S)$$

$$= \frac{2N(N+1)}{12}(1 - r_S)$$

so that

$$r_S = 1 - \frac{6S_d^2}{N(N+1)}$$

By definition,

$$S_d^2 = \frac{N\left(\sum_{i=1}^{N} d_i^2\right) - \left(\sum_{i=1}^{N} d_i\right)^2}{N(N-1)}$$

$$= \frac{N\sum_{i=1}^{N}(R_{1i} - R_{2i})^2 - \left[\sum_{i=1}^{N}(R_{1i} - R_{2i})\right]^2}{N(N-1)}$$

Since

$$\sum_{i=1}^{N} (R_{1i} - R_{2i}) = \sum_{i=1}^{N} R_{1i} - \sum_{i=1}^{N} R_{2i}$$

$$= \frac{N(N+1)}{2} - \frac{N(N+1)}{2}$$

$$= 0$$

it follows that

$$S_d^2 = \frac{N \sum_{i=1}^{N} (R_{1i} - R_{2i})^2}{N(N-1)} = \frac{\sum_{i=1}^{N} (R_{1i} - R_{2i})^2}{(N-1)}$$

so that

$$r_S = 1 - \frac{6 \sum_{i=1}^{N} (R_{1i} - R_{2i})^2}{N(N-1)(N+1)}$$

$$= 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N^3 - N}$$

This completes the proof.

Summary statistics for the determination of r_S for the AP–UP ratings of football teams are shown in Table 19-2. For these data,

$$\sum_{i=1}^{17} (R_{1i} - R_{2i})^2 = 0^2 + 0^2 + \cdots + (-1)^2 + (1)^2 + \cdots + (-4)^2 + \cdots + (4)^2 + 0^2$$

$$= 34$$

and

$$r_S = 1 - \frac{6(34)}{17^3 - 17} = .958$$

Since r_S is so close to 1, it must be concluded that both polls are measuring the same elements in football performance.

When $N \geqslant 10$, one can test the hypothesis H_0: $E(r_S) = 0$ versus H_1: H_0 is false by means of

$$t = \frac{\sqrt{N-2}\, r_S}{\sqrt{1 - r_S^2}}$$

Table 19-2. Summary statistics for the determination of the Spearman rank correlation coefficient for the AP–UP football ratings of 20 November 1968.

School	*No of votes* AP	UP	*Rank* AP	UP	*Difference*
Southern California	704	338	1	1	0
Ohio State	636	309	2	2	0
Penn State	571	280	3	3	0
Michigan	545	207	4	4	0
Georgia	530	192	5	5	0
Texas	359	155	6	7	−1
Kansas	330	156	7	6	+1
Tennessee	256	80	8	8	0
Arkansas	236	55	9	9	0
Notre Dame	205	52	10	10	0
Houston	120	31	11	11	0
Purdue	118	8	12	16	−4
Missouri	110	15	13	13	0
Oklahoma	92	13	14	14	0
Alabama	65	9	15	15	0
Oregon State	56	16	16	12	+4
Ohio University	41	4	17	17	0

which has an approximate t distribution with $\nu = N - 2$ when the hypothesis being tested is true. For the football poll data,

$$t = \frac{\sqrt{17-2}(.958)}{\sqrt{1-(.958)^2}} = 12.93$$

which, with $\nu = 15$, indicates that $E(r_S) \neq 0$.

19-5 REPEATED MEASURES—THE EXTENSION OF THE MATCH PAIR t TEST

In Section 11-10, the match pair t test was presented for testing H_0: $\mu_1 = \mu_2$ versus H_1: H_0 is false. Many experimental studies in psychology and education extend the model of the match pair t test one step further by instituting K repeated measures on each subject instead of the two measures that are involved in the match pair t test. This extension is illustrated in Table 19-3, which shows the number of correct responses by eight subjects placed in a control group of a larger serial learning study done by Levin in 1968 (unpublished dissertation).

Table 19-3. Number of correct responses made by eight subjects placed in a control condition under the method of serial anticipation (maximum score is equal to 12).

Subject	*Trial*					*Total*
	1	2	3	4	5	
1	0	3	2	3	4	12
2	6	8	9	9	9	41
3	9	9	10	12	11	51
4	5	6	6	6	7	30
5	6	8	5	11	11	41
6	3	3	6	7	10	29
7	4	5	6	7	6	28
8	3	6	9	8	10	36
Total	36	48	53	63	68	268
Mean	4 50	6.00	6 625	7 875	8 50	6 70
S of *S*	212	324	399	553	624	2112

As the arrangement of data in Table 19-3 suggests, the hypothesis of interest is given by $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ versus H_1: H_0 is false. In terms of the analysis-of-variance model, one might think that the appropriate test statistic is given by $F = \text{MSB}/\text{MSW}$. In this case, $T_{.1} = 36$, $T_{.2} = 48$, $T_{.3} = 53$, $T_{.4} = 63$, $T_{.5} = 68$, and $T_{..} = 268$, so that

$$\text{I} = \sum_{k=1}^{5} \sum_{i=1}^{8} y_{ik}^2 = 0^2 + 6^2 + 9^2 + \cdots + 10^2 + 6^2 + 10^2 = 2112$$

$$\text{II} = \frac{T^2}{NK} = \frac{1}{(8)(5)}(268)^2 = 1{,}795.60$$

$$\text{III}_C = \sum_{k=1}^{K} \frac{T_{.k}^2}{N} = \tfrac{1}{8}[36^2 + 48^2 + 53^2 + 63^2 + 68^2] = 1{,}875.25$$

$$\text{SSB} = \text{III}_C - \text{II} = 1{,}875.25 - 1{,}795.60 = 79.65$$

$$\text{SSW} = \text{I} - \text{III}_C = 2{,}112 - 1{,}875.25 = 236.75$$

$$\text{SST} = \text{SSB} + \text{SSW} = 79.65 + 236.75 = 316.40$$

$$\text{MSB} = \frac{1}{K-1}\text{SSB} = \tfrac{1}{4}(79.65) = 19.91$$

$$\text{MSW} = \frac{1}{K(N-1)}\,\text{SSW} = \frac{1}{(5)(7)}(236.75) = 6.76$$

$$F = \frac{\text{MSB}}{\text{MSW}} = \frac{(19.91)}{(6.76)} = 2.94$$

With $\alpha = .01$, $F_{4,35}(.99) = 3.92$, so that H_0 is not rejected.

This analysis is valid if the underlying variable is normal or if the individual sample sizes are large enough so that the sampling distribution of the means is near normal in form. In addition, it must be assumed that observations within a trial period are independent. In this case, these assumptions are not suspect. However, the assumption that observations between trials are independent is suspect. For example, the scores made by subject 1 are low in the beginning and remain low at the end. On the other hand, subject 3 scores high on trial 1 and also scores high on trial 5. Without doubt, the scores across trials by subject are correlated. Thus, the analysis-of-variance model, as presented, is not valid. Fortunately, the adjustment for correlated or repeated measures, as the model is usually referred to, is easy to make. For this adjustment, consider an analysis of variance on the *subjects*. For the eight subjects, one has $T_1 = 12$, $T_2 = 41$, $T_{3.} = 51$, $T_4 = 30$, $T_{5.} = 41$, $T_{6.} = 29$, $T_{7.} = 28$, $T_8 = 36$, so that

$$\text{III}_S = \sum_{i=1}^{8} \frac{T_i^2}{K} = \tfrac{1}{5}[12^2 + 41^2 + 51^2 + 30^2 + 41^2 + 29^2 + 28^2 + 36^2] = 1{,}985.60$$

$$\text{SSS} = \text{III}_S - \text{II} = 1{,}985.60 - 1{,}795.60 = 190.00$$

$$\text{MSS} = \frac{1}{N-1}\text{SSS} = \tfrac{1}{7}(190.00) = 27.14$$

It may come as a surprise that the SSS = 190.00 is included in the SSW = 236.75, but a little reflection will show that this is the case. From the model presented in Chapter 15, and from this last observation, it is reasonable to assume that

$$\text{SST} = \text{SSB} + \text{SSS} + \text{SSR}$$

where SSR = sum of squares residual. If one substitutes the known sum of squares in this equation, it is seen that

$$316.40 = 79.65 + 190.00 + \text{SSR}$$

so that

$$\text{SSR} = 46.75$$

It should be noted that of the total SSW = 236.75, most was attributable to SSS = 190.00. In psychological and educational studies it is not unusual to find that most of the variability in the criterion measures is directly traceable to the individual differences that exist between human subjects.

From the analysis-of-variance model of Chapter 15, one would expect that

$$\nu_T = \nu_B + \nu_S + \nu_R$$

or that

$$(KN - 1) = (K - 1) + (N - 1) + \nu_R$$

so that

$$\begin{aligned}\nu_R &= (KN - 1) - (K - 1) - (N - 1)\\ &= KN - K - N + 1\\ &= (K - 1)(N - 1)\end{aligned}$$

For this study,

$$\nu_R = (5 - 1)(8 - 1) = (4)(7) = 28$$

The complete analysis-of-variance table for the study is shown in Table 19-4. As can be seen, the expected mean squares involve σ_S, the variance between subjects

Table 19-4. Analysis-of-variance table for the repeated-measures study of Table 19-3.

Source	*d/f*	*S* of *S*	MS	*F*	*E*(MS)
Between trials	4	79 65	19.91	11.92	$\sigma^2(1-\rho) + \frac{N}{K-1}\sum_{k=1}^{K}(\mu_k - \mu)^2$
Between subjects	7	190.00	27.14	16.75	$\sigma^2(1-\rho) + K\sigma_S^2$
Residual	28	46.75	1.67		$\sigma^2(1-\rho)$
Total	39	316 40			

and ρ, the correlation between the scores on the trials. Unfortunately, the derivation of these expected values is not easy and no useful purpose would be satisfied in deriving them in this book. For this reason they are stated without proof.

Theorem 19-8

When the variance between subjects is not zero and when there are treatment effects in a repeated-measures design, then

$$E(\text{MSB}) = \sigma^2(1-\rho) + \frac{N}{K-1}\sum_{k=1}^{K}(\mu_k - \mu)^2$$

$$E(\text{MSS}) = \sigma^2(1-\rho) + K\sigma_S^2$$

$$E(\text{MSR}) = \sigma^2(1-\rho)$$

where

$$\sigma^2 = \text{variance within treatments}$$
$$(\mu_k - \mu) = k\text{th treatment effect}$$
$$\sigma_S^2 = \text{variance between subjects}$$
$$\rho = \text{correlation coefficient between treatments}$$

As might be expected, the proof that $E(\text{MSR}) = \sigma^2(1 - \rho)$ is related to Theorem 19-1, which is also used in obtaining the test statistic for the match pair t test.

Since MSB is an estimate of

$$\sigma^2(1 - \rho) + \frac{N}{K-1} \sum_{k=1}^{K} (\mu_k - \mu)^2$$

and since MSR is an estimate of $\sigma^2(1 - \rho)$, it then follows that if $F = \text{MSB}/\text{MSR}$ is larger than could be expected on the basis of chance, it must follow that

$$\frac{N}{K-1} \sum_{k=1}^{K} (\mu_k - \mu)^2 > 0$$

When this occurs, then H_0 must be false. In this sense, the effect of ρ upon the test statistic is quite apparent, since $E(\text{MSW}) < E(\text{MSR})$. When $\rho = 0$, as it is in the analysis-of-variance design of Chapter 15, then $E(\text{MSW}) = E(\text{MSR})$. Since $F_{4,28}(.99) = 4.07$, the hypothesis of equal mean scores across trials is rejected, since $F = 11.92$. This is not the same conclusion reached when the F test was performed ignoring the correlation between trials. When the correlation of the reported study was ignored, F was given by $F = 2.94$, while when the correlation was considered, F was given by $F = 11.92$. This suggests that the power of the test has been substantially increased. This increase in power is directly related to the sample estimate of the correlations. As ρ approaches 1, the power of the test also approaches 1. In the sample, the unbiased estimate of $E(\text{MSR}) = \sigma^2(1 - \rho)$ is given by

$$\text{MSR} = \text{MSW}(1 - \bar{r})$$

For the observed data,

$$1.67 = 6.76(1 - \bar{r})$$

so that

$$\bar{r} = \tfrac{5.09}{6.76} = .76$$

It should be noted that the assumptions of the analysis-of-variance model of Chapter 15 are still required for the valid use of this test. In addition, it must be assumed that all correlations between trials are identical. For the example of this section, it means that $\rho_{12} = \rho_{13} = \rho_{14} = \rho_{15} = \rho_{23} = \rho_{24} = \rho_{25} = \rho_{34} = \rho_{35} = \rho_{45}$. The sample correlation $\bar{r} = .76$ is the estimate of the average of these correlations. While

the test is based on the equality of the $\binom{K}{2}$ correlations, one can tolerate moderate departures from equality because of the robustness of the F test. However, larger variabilities in the sample correlation coefficients should be looked upon as deviations from the test assumptions.

Finally, by using Theorem 19-9 and Scheffé's theorem, we can perform *post hoc* comparisons upon the sample averages, although the MSR is used for computing the standard error of the contrasts.

Theorem 19-9

The squared standard error for the difference between two sample means in a repeated-measures design is given by

$$\mathrm{SE}^2_{(\bar{Y}_{k_1}-\bar{Y}_{k_2})} = \frac{2}{N}\mathrm{MSR}$$

Proof. According to Theorem 19-1,

$$\mathrm{SE}^2_{(\bar{Y}_{k_1}-\bar{Y}_{k_2})} = S^2_{\bar{Y}_{k_1}} + S^2_{\bar{Y}_{k_2}} - 2\bar{r}S_{\bar{Y}_{k_1}}S_{\bar{Y}_{k_2}}$$

but since the variances, correlations, and sample sizes are all equal,

$$\begin{aligned}
\mathrm{SE}^2_{(\bar{Y}_{k_1}-\bar{Y}_{k_2})} &= \frac{\mathrm{MSW}}{N} + \frac{\mathrm{MSW}}{N} - 2\bar{r}\frac{\sqrt{\mathrm{MSW}}}{\sqrt{N}}\frac{\sqrt{\mathrm{MSW}}}{\sqrt{N}} \\
&= \left(\frac{\mathrm{MSW}}{N} - \bar{r}\frac{\mathrm{MSW}}{N}\right) + \left(\frac{\mathrm{MSW}}{N} - \bar{r}\frac{\mathrm{MSW}}{N}\right) \\
&= \frac{\mathrm{MSW}(1-\bar{r})}{N} + \frac{\mathrm{MSW}(1-\bar{r})}{N} \\
&= \frac{1}{N}\mathrm{MSR} + \frac{1}{N}\mathrm{MSR} \\
&= \frac{2}{N}\mathrm{MSR}
\end{aligned}$$

This completes the proof.

Thus, one need not worry about the correlation. Once the SSS is removed from SSW, the correction for correlation is achieved.

For completeness, Table 19-5 is presented to illustrate the general form of the analysis of variance for a repeated-measures study based on N subjects and K experimental conditions.

Table 19-5. The general analysis-of-variance table for a repeated-measures study based on N subjects and K experimental conditions.

Source of variance	*d/f*	*S* of *S*	MS	*F*	*E*(MS)
Between conditions	$K-1$	$\text{III}_B - \text{II}$	MSB	$\frac{\text{MSB}}{\text{MSR}}$	$\sigma^2(1-\rho) + \frac{N}{K-1}\sum_{i=1}^{N}(\mu_k - \mu)^2$
Between subjects	$N-1$	$\text{III}_S - \text{II}$	MSS	$\frac{\text{MSS}}{\text{MSR}}$	$\sigma^2(1-\rho) + K\sigma_S^2$
Residual	$(K-1)(N-1)$	$\text{I} - \text{III}_B - \text{III}_S + \text{II}$	MSR		$\sigma^2(1-\rho)$
Total	$KN-1$	$\text{I} - \text{II}$	MST		

19-6 COCHRAN'S Q

A special case of the repeated-measures design occurs when the dependent variable is binomial. When this occurs, the test of hypothesis of equal expected values is identical to the test of equal binomial probabilities. Thus, if X is $\{0,1\}$, then the test of equal expected values reduces to $H_0: p_1 = p_2 = \cdots = p_K = p_0$ versus H_1: H_0 is false. While this hypothesis is identical to the hypothesis of equal p values discussed in Section 16-2, the test of that section is invalid because the observations within a subject are correlated A test of hypothesis of equal correlated proportions was derived by Cochran and is generally denoted as Cochran's Q. The test statistic for this test is given by

$$Q = \frac{K(K-1)\sum_{k=1}^{K}(T_{.k} - T_{.})^2}{\sum_{i=1}^{N} T_{i.} - \left(\sum_{i=1}^{N} T_i\right)^2}$$

where $T_{.k}$ = total number of 1's in condition k
T_i = total number of 1's for ith subject
$T_.$ = average number of 1's in the K conditions
N = number of subjects

If N is sufficiently large, then Q is approximately distributed as a chi-square variable with $\nu = K - 1$. If H_0 is rejected, one can locate the source of rejection by placing confidence intervals about contrasts of the form

$$\psi = a_1 p_1 + a_2 p_2 + \cdots + a_K p_K$$

for which the squared standard error is given by

$$\mathrm{SE}_{\hat{\psi}}^2 = \left(\frac{K \sum_{i=1}^{N} T_i - \sum_{i=1}^{N} T_{i.}^2}{NK(K-1)} \right) \sum_{k=1}^{K} \frac{a_k^2}{N}$$

and for which

$$S^* = \sqrt{\chi_{K-1}^2(1-\alpha)}$$

If $K = 2$, the test is called McNemar's Test for Equality of Correlated Proportions.

It should be noted that the use of this test and confidence-interval procedures requires that N be large and that no subject give consistent responses across all trials.

19-7 STATISTICAL TESTS AND CONFIDENCE INTERVALS FOR THE BIVARIATE NORMAL CORRELATION MODEL

Table 19-6 provides a summary of the statistical tests and procedures presented in this chapter.

19-8 SUMMARY

In repeated samples from $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, the sampling distribution of r is very complex. A transformation of the distribution of r to an approximately normal form was derived by R. A. Fisher. This transformation is given by

$$z = \frac{1}{2} \log_e \frac{1+r}{1-r}$$

which produces a near-normal variable such that

$$E(z) = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho}$$

and

$$\mathrm{Var}(z) = \frac{1}{N-3}$$

This transformation can be used to test hypotheses and construct confidence intervals about ρ. To test H_0: $\rho = \rho_0$ versus H_1: H_0 is false, the appropriate test statistic is given by

$$Z = \frac{z - E(z)}{\sigma_z}$$

Table 19-6. Tests and confidence intervals for quantitative correlation models.

Case	*Hypothesis*	*Test statistic*	*Confidence interval*	*Assumptions*
19	$H_0: \rho = \rho_0$ $H_1: \rho \neq \rho_0$	$Z = \dfrac{z - E(z)}{\sigma_z}$ where $z = \frac{1}{2}\log_e \dfrac{1+r}{1-r}$ $\sigma_z = \dfrac{1}{\sqrt{N-3}}$	$z - Z\left(\frac{\alpha}{2}\right)\sigma_z < E(z) < z + Z\left(\frac{\alpha}{2}\right)\sigma_z$	1. Y_1 and Y_2 are bivariate normal 2 Independence between pairs of observations
20	H_0 $\rho_1 = \rho_2$, or $\theta = \rho_1 - \rho_2 = 0$ $H_1: H_0$ is false	$Z = \dfrac{(z_1 - z_2)}{\sqrt{1/(N_1-3) + 1/(N_2-3)}}$	$E(z_1) - E(z_2) = (z_1 - z_2)$ $\pm Z\left(\frac{\alpha}{2}\right)\sqrt{\dfrac{1}{N_1-3} + \dfrac{1}{N_2-3}}$	1 Y_1 and Y_2 are bivariate normal in each universe 2. Independence between pairs of observations 3. Independence between samples
21	$H_0: \rho_1 = \rho_2 = \cdot\ \cdot = \rho_K$ $H_1: H_0$ is false	$U_0 = \sum_{k=1}^{K} (N_k - 3)(z_k - z_0)^2$ where $\nu = K - 1$ $z_k = \frac{1}{2}\log_e \dfrac{1+r_k}{1-r_k}$ $z_0 = \dfrac{\sum_{k=1}^{K}(N_k - 3)\, z_k}{\sum_{k=1}^{K}(N_k - 3)}$	$E(z_{k_1}) - E(z_{k_2}) = (z_{k_1} - z_{k_2})$ $\pm \sqrt{\chi^2_{K-1}(1-\alpha)}\sqrt{\dfrac{1}{N_{k_1}-3} + \dfrac{1}{N_{k_2}-3}}$ $k_1, k_2 = 1, 2, \ldots, K$ $k_1 < k_2$	1 Y_1 and Y_2 are bivariate normal in each universe 2 Independence between pairs of observations 3. Independence between samples

which, when H_0 is true, is $N(0,1)$. To test H_0: $\rho_1 = \rho_2$ versus H_1: H_0 is false, the appropriate test statistic is given by

$$Z = \frac{z_1 - z_2}{\sqrt{1/(N_1 - 3) + 1/(N_2 - 3)}}$$

which, when H_0 is true, is $N(0,1)$ Finally, to test H_0: $\rho_1 = \rho_2 = \cdots = \rho_K$ versus H_1: H_0 is false, the appropriate test statistic is given by

$$U_0 = \sum_{k=1}^{K} (N_k - 3)(z_k - z_0)^2$$

where

$$z_0 = \frac{\sum_{k=1}^{K} (N_k - 3) z_k}{\sum_{k=1}^{K} (N_k - 3)}$$

which, when H_0 is true, is χ^2_{K-1}.

For the one-sample problem, $(1 - \alpha)$ percent confidence intervals for ρ are given by retransforming the confidence interval

$$z - Z\left(\frac{\alpha}{2}\right)\sigma_z < E(z) < z + Z\left(\frac{\alpha}{2}\right)\sigma_z$$

into one for ρ. In the K-sample problem, corresponding confidence intervals are given by

$$\psi = (z_{k_1} - z_{k_2}) \pm \sqrt{\chi^2_{K-1}(1-\alpha)}\sqrt{\frac{1}{N_{k_1} - 3} + \frac{1}{N_{k_2} - 3}}$$

with $k_1 < k_2$ and $k = 1, 2, \ldots, K$. In this case, one does not transform back to the parameters of the distributions. If a difference in z values does not include zero, it is concluded that the correlations are different.

When data are ranked, one can use Spearman's rank correlation coefficient

$$r_S = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N^3 - N}$$

to estimate the strength of association between two variables that are believed to be related. In this formula, $d_i = R_{1i} - R_{2i}$ represents the difference in rank for the ordered observations.

The appearance of correlated observations in educational and psychological investigations is a common occurrence within the framework of an analysis-of-variance design. If K repeated measures are taken on N different subjects and if

the outcomes for the ith subject are denoted by $(Y_{1i}, Y_{2i}, \ldots, Y_{Ki})$, then the appropriate analysis-of-variance table for testing H_0: $\mu_1 = \mu_2 = \cdots = \mu_K$ versus H_1: H_0 is false is as shown in Table 19-5 and the appropriate test statistic is given by:

$$F = \frac{\text{MSB}}{\text{MSR}} = \frac{\text{SSB}/(K-1)}{\text{SSR}/(N-1)(K-1)}$$

where

$$\text{SSB} = \text{III}_B - \text{II} = \sum_{k=1}^{K} \frac{T_{.k}^2}{N} - \frac{T_{..}^2}{NK}$$

and

$$\text{SSR} = \text{I} - \text{III}_B - \text{III}_S + \text{II}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} y_{ik}^2 - \sum_{k=1}^{K} \frac{T_{.k}^2}{N} - \sum_{i=1}^{N} \frac{T_{i.}^2}{K} + \frac{T_{..}^2}{NK}$$

The assumptions for this test are:

1. Independence between the subjects
2. Equal variances for the study conditions
3. Normality of the parent populations
4. Equal correlations between conditions

This last assumption is sometimes difficult to satisfy.

If H_0 is rejected, one can locate the source of the variance by simple application of Scheffé's theorem to contrasts involving mean differences. For the estimated contrast $\hat{\psi} = \bar{Y}_{k_1} - \bar{Y}_{k_2}$, the $(1-\alpha)$ percent confidence interval for $\psi = \mu_{k_1} - \mu_{k_2}$ is given by

$$(\bar{Y}_{k_1} - \bar{Y}_{k_2}) \pm \sqrt{(K-1) F_{(K-1),(N-1)(K-1)}(1-\alpha)}\, \text{SE}_{\hat{\psi}}$$

where

$$\text{SE}_{\hat{\psi}}^2 = \text{SE}_{\bar{Y}_{k_1}}^2 - 2\bar{r}\text{SE}_{\bar{Y}_{k_1}} \text{SE}_{\bar{Y}_{k_2}} + \text{SE}_{\bar{Y}_{k_2}}^2$$

Since

$$\text{SE}_{\bar{Y}_{k_1}}^2 = \text{SE}_{\bar{Y}_{k_2}}^2 = \frac{\text{MSW}}{N}$$

$$\text{SE}_{\hat{\psi}}^2 = \frac{\text{MSW}}{N} - 2\bar{r}\frac{\text{MSW}}{N} + \frac{\text{MSW}}{N}$$

$$= \frac{\text{MSW}}{N}(1-\bar{r}) + \frac{\text{MSW}}{N}(1-\bar{r})$$

$$= \frac{\text{MSR}}{N} + \frac{\text{MSR}}{N} = \frac{2}{N}\text{MSR}$$

one need not compute $\bar{r}$ to use the Scheffé method. However, if the value of the common sample correlation coefficient is desired it can be simply computed from

$$\text{MSR} = \text{MSW}(1 - \bar{r})$$

which gives

$$\bar{r} = \frac{\text{MSW} - \text{MSR}}{\text{MSW}}$$

When the underlying variable in a repeated-measures design is binomial, the test of equal expected values reduces to $H_0: p_1 = p_2 = \cdots = p_K = p_0$ versus H_1: H_0 is false, with the test statistic given by

$$Q = \frac{K(K-1)\sum_{k=1}^{K}(T_{.k} - T_{.})^2}{K\sum_{i=1}^{N} T_{i.} - \sum_{i=1}^{N} T_{i.}^2}$$

A simple computing formula for Q is given by

$$Q = \frac{(K-1)\left[K\sum_{k=1}^{K} T_{.k}^2 - \left(\sum_{k=1}^{K} T_{.k}\right)^2\right]}{K\sum_{i=1}^{N} T_{i.} - \sum_{i=1}^{N} T_{i.}^2}$$

If H_0 is rejected, *post hoc* confidence intervals may be investigated with $\hat{\psi} = a_1\hat{p}_1 + a_2\hat{p}_2 + \cdots + a_K\hat{p}_K$, and

$$\text{SE}_{\hat{\psi}}^2 = \left[\frac{K\sum_{i=1}^{N} T_{i.} - \sum_{i=1}^{N} T_{i.}^2}{NK(K-1)}\right]\sum_{k=1}^{K}\frac{a_k^2}{N}$$

and

$$S^* = \sqrt{\chi_{K-1}^2(1-\alpha)}$$

EXERCISES

***19-1.** For the data of Table 18-3, is there any evidence that $\rho = 0$?

***19-2.** For the data of Table 18-3, is there any evidence that $\rho \geqslant .8$?

***19-3.** For the data of Table 18-3, determine the 95 percent confidence interval for μ_B using the models of Sections 19-2 and 11-4. Which do you prefer? Why?

***19-4.** For the data of Table 18-3, determine the 95 percent confidence interval for $\sigma_{B\,G}^2$. How does it relate to σ_B^2?

19-5. In a study in which the effects that the number of periods of training had upon the learning of a serial list of nonsense words, each subject was scored as having learned the

list (1) or not learned the list (0) The results of the testing after each training period is as shown in the following table Use Cochran's Q to assess the outcomes of this study.

Subject	*Number of training periods* 1	2	3	4
Alice	0	0	1	1
Betty	0	1	0	1
Carol	0	0	1	1
Donna	1	1	0	1
Elvira	1	1	0	1
Francine	0	0	1	1
Grace	1	1	0	0
Hilda	0	0	1	1
Ida	1	0	0	1
Janet	0	0	1	1

***19-6.** For the data of Table 5-8,

(*a*) Show that

$$S_d^2 = S_N^2 + S_T^2 - 2rS_N S_T$$

(*b*) Test the hypothesis of no difference in expected reading scores for the two experimental treatments.
(*c*) What have you assumed?
(*d*) Are the assumptions reasonable? Explain.

***19-7.** (*a*) Make the scatter plot for the data of Table 13-4.
(*b*) Do the variables appear to be bivariate normal? Why?
(*c*) Estimate the regression of Y_1 on Y_2 and Y_2 on Y_1.
(*d*) Estimate $E(Y_1 | Y_2 = 60)$ and $E(Y_2 | Y_1 = 60)$. Why are they different from one another?
(*e*) Set up a 95 percent confidence interval for $E(Y_2)$.
(*f*) Are the two variables independent of one another?
(*g*) Determine the Spearman rank correlation coefficient for this data How does it compare to r?

***19-8.** In Exercise 19-7, an earlier study made in 1960 showed that $\rho = .6$ Has the correlation changed between 1960 and 1965? What have you assumed in answering this question?

19-9. In a comparative study of the association between internal anxiety, as measured by a galvanic skin response when subjects were placed under psychological stress, and a

tendency to exhibit hysterical behavior, as measured by a questionnaire inventory, the following statistics were generated.

Group	*Sample size*	*Correlation coefficient*	*Average age*
College sophomores	33	26	18.6
Manic depressive institutionalized patients	21	.59	32.7
Army recruits	36	.40	19.1
Campfire girls	19	.51	12 7

On the basis of these data, can it be concluded that the degree of the association between the test variables is independent of the population tested? Can one generalize to other populations such as housewives, union members, Catholics, school teachers, etc? Defend your answer.

***19-10.** Perform the Scheffé contrasts for the data of Table 19-3 to locate the possible sources of variance. What are the findings? How has the correlation coefficient entered into the variances of the contrasts?

20 INTRODUCTION TO STATISTICAL REGRESSION THEORY

Which aspect of the influence of a majority is more important—the size of the majority or its unanimity? . When a subject was confronted with only a single individual who contradicted his answers, he was swayed little he continued to answer independently and correctly in nearly all trials. When the opposition was increased to two, the pressure became substantial: minority subjects now accepted the wrong answer 13.6 percent of the time. Under the pressure of a majority of three, the subjects' errors jumped to 31.8 percent But further increases in the size of the majority apparently did not increase the weight of the pressure substantially Clearly the size of the opposition is important only up to a point.

20-1 CHARACTERISTICS OF REGRESSION STUDIES

Many studies that pass in the applied and research literature of the behavioral sciences as correlational studies are in reality regression studies. To many behavioral scientists, this statement is void of meaning, since many believe that a correlational study is equivalent to a regression study. On the surface this supposition carries some truth; yet, the two kinds of studies provide answers to different questions. Confusion between the two statistical models is easy to understand, but once the distinction between them is understood, the incorrect application of one model for the other should be minimized.

While a correlation problem involves one sample from one bivariate universe, a regression problem typically involves K univariate samples selected from K independent universes, which are defined in terms of a quantitative variable While a correlation problem involves the joint observation of two random variables, a regression problem is concerned with observing one random variable for fixed values of another variable where the exact numerical values of the second variable are specified in advance of data collection. Since a regression problem is a K sample univariate problem in which the variable of interest is continuous or discrete, it might be reasonable to assume that analysis-of-variance methods could be used to study this problem. While this is indeed true, this procedure is not emphasized in this book, but for completeness is illustrated in Section 20-10.

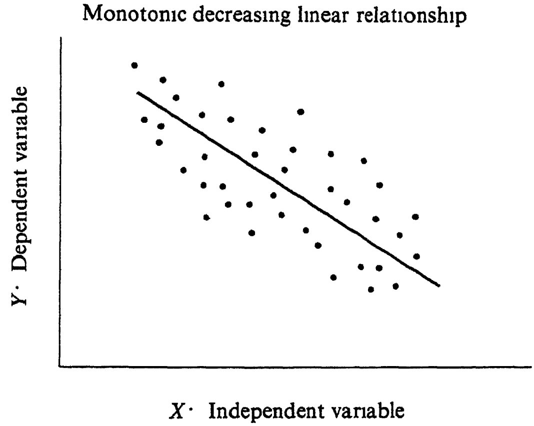
In classical correlation theory, Y_1 and Y_2 have equal importance. Both are random variables with respective probability distributions. In a regression study, a different probability model is assumed. One of the variables, X, is not a random variable but is in reality a parameter variable in direct control by the researcher. Its values are fixed before the study is begun. Such a manipulated variable is popularly referred to as the *independent variable* of the study. For each fixed value of the independent variable X, another variable Y, called the *dependent variable*, is observed.

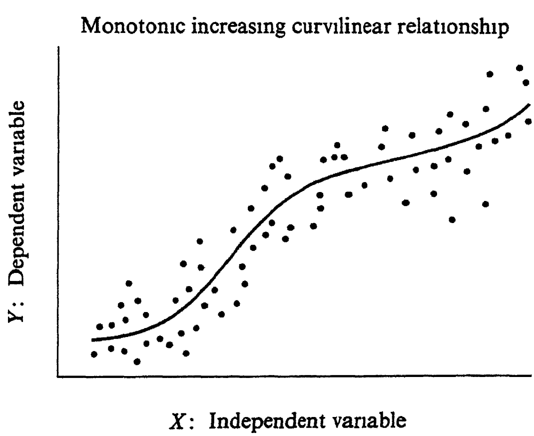
While classical correlation theory is based on a linear relationship between Y_1 and Y_2, classical regression theory is not so limited. This is of considerable value since curvilinear relationships are commonly observed in behavioral research studies. A common finding of educational and psychological research is that the relationship between the independent variable X and the dependent variable Y is *monotonic*. By this is meant that as X increases in value, Y always increases or always decreases. If Y always increases as X increases, it is said that the relationship is monotonic increasing, if, on the other hand, Y always decreases as X increases, it is said that the relationship is monotonic decreasing. Examples of monotonic relationships are shown in Figure 20-1.

The distinction between independent and dependent variables has been extended to include models not customarily treated as regression models and in which "nature" may have decided upon the characteristics of the independent variable. For example, in a study in which the effects of race upon school achievement is being assessed, the independent variable of the study, race, is usually genetically determined by nature, but so as to help understand what the data show about school achievement,

a researcher may define the independent variable by the following set of mutually exclusive and exhaustive subsets {Caucasian, Negro, Oriental, other} The dependent variable of such a study might be the score achieved on some paper-and-pencil performance test so that the independent variable is qualitative while the dependent variable is discrete. Other examples could be presented without limit As an exercise, it is informative to consider each example of Chapter 11 and to identify the independent variables of each study presented.

Figure 20-1. Examples of a monotonic decreasing linear relationship and a monotonic increasing curvilinear relationship.





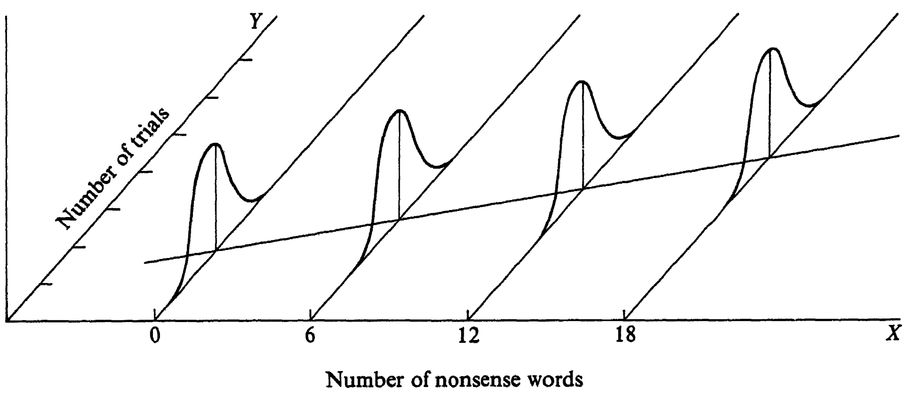
20-2 EXAMPLE OF A REGRESSION PROBLEM

To gain some insight into a regression problem and to see how it differs from a correlation problem, consider the following hypothetical experimental situation. Thirty-two students are to be divided at random into four groups and the members of each group are going to be asked to memorize a five-line poem containing nonsense words. The four experimental conditions of the study are defined in terms of the number of nonsense words placed in the poem by the experimenter. For the eight control subjects in condition 1 none of the nouns are replaced by nonsense words; for the students in condition 2, six of the nouns are replaced with six different nonsense words. For the eight subjects in condition 3, twelve of the nouns are replaced by nonsense words, and for the remaining eight subjects in condition 4, eighteen of the nouns are replaced by nonsense words. Under these conditions, it is reasonable to expect the number of trials it takes to learn the poem to increase as the number of nonsense words increases. Thus, an increasing monotonic relationship between the expected values of the four conditions is anticipated. If the relationship were linear, the expected values for each of the four conditions would lie on a straight line, as shown in Figure 20-2. The independent or predetermined fixed variable for this study is the number of nonsense words in the poem, while the observed or dependent variable is the number of trials it takes to learn the poem.

It is worth noting that this model can in no way be construed to represent a joint bivariate normal model. This study involves four unique distributions defined at $X = 0, 6, 12, 18$ and is clearly in the mode of an analysis-of-variance study.

Returning to the experimental study, it seems reasonable to assume or hypothesize that as new nonsense words are added to the poem the number of trials it takes to learn the poem will increase in a fixed manner. For example, if it takes 10 trials to learn the poem with 0 nonsense words and 12 extra trials to learn the poem with 6 nonsense words, then it might be reasonable to assume that a poem with 12 nonsense words should take $10 + 2(12) = 34$ trials, and a poem with 18 nonsense words should

Figure 20-2. Three-dimensional representation of a linear regression model for $K = 4$



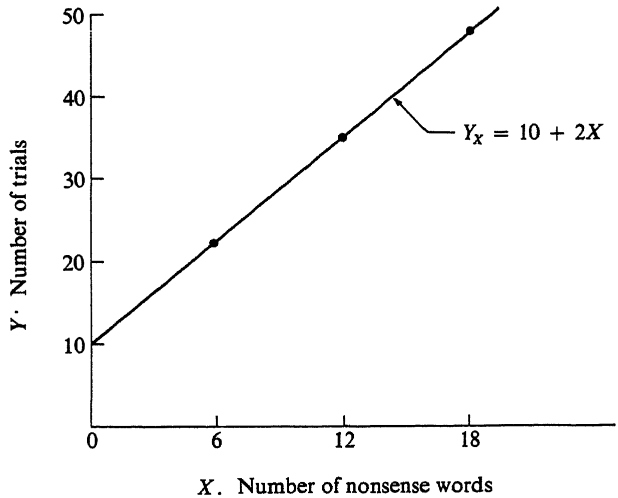


Figure 20-3. Hypothetical linear relationship between number of trials to learn a poem and number of nonsense words contained in the poem.

take $10 + 3(12) = 46$ trials. Essentially this means that the expected values are expected to increase in a linear manner. That is, if the average number of trials is graphed against the number of nonsense words and if the plotted points are joined by line segments, the line segments will in reality trace a straight line, as shown in Figure 20-3.

When the population means can be joined by a single straight line, it is said that the means satisfy a linear regression equation. In this example the equation of the regression line is given by

$$Y_X = 10 + 2X$$

The intercept of the regression equation is equal to 10. It equals the mean value of the dependent variable when the independent variable is 0. The slope of the regression equation is equal to 2. It specifies the average increase in the dependent measure for each unit increase in the independent measure. When $X = 12$, $Y_{12} = 10 + 2(12) = 34$ or when $X = 13$, $Y_{13} = 10 + 2(13) = 36$, so that $Y_{13} - Y_{12} = 36 - 34 = 2$, the average increase in Y for each unit increase in X.

As this discussion indicates, a regression model assumes the existence of an independent variable or a variable that behavioral scientists can manipulate and a dependent or criterion variable that is observed on each level of the independent variable. This differs from the correlation model in that for the correlation model there are no clear-cut dependent and independent variables. Both variables covary together so that two different regression lines are generated. For one of the regression lines, Y_1 is treated as though it were an independent variable and in the other

regression line, Y_2 is treated as though it were an independent variable. Of course, neither are independent variables and neither are dependent variables. Finally, it should be noted that a regression problem involves only one regression equation.

Sometimes regression models are called prediction models The reason for this is easy to understand. Suppose that the regression equation is given by $Y_X = 10 + 2X$. If X is known to be equal to 14, then the predicted value of Y is given by $Y_{14} = 10 + 2(14) = 38$ Of course, in any one case the observed value of Y may not be equal to 38. If the association between X and Y is strong, then the deviation from prediction will be small.

The results for the 32 subjects of the experiment are listed in Table 20-1. It appears that the number of trials to learn the poem increases as the number of nonsense

Table 20-1. Number of trials to learn the poem for the four experimental conditions of the study.

Student number in group	*Number of nonsense words in poem* 0	6	12	18
1	4	9	7	15
2	8	8	15	18
3	7	12	12	20
4	6	10	9	14
5	10	7	16	12
6	6	13	14	19
7	3	9	13	19
8	5	10	15	16
Average	6 125	9.750	12.625	16.625

words increases. The conditional sample mean of each treatment group is reported in the table. These means are plotted in Figure 20-4. As this figure suggests, the unknown expected values appear to lie on a straight line. If the equation of the line were known, it could be used to predict the expected values for other values of Y. As will be seen, the estimates of the conditional expected values based upon the regression line are more efficient estimates than are the simple averages of the number of trials of 0, 6, 12, and 18 nonsense words.

As the data are presented, it appears that one could perform an analysis of variance on the data and test H_0: $\mu_0 = \mu_6 = \mu_{12} = \mu_{18}$ versus the alternative H_1: H_0 is false. While this type of analysis is not incorrect, it is not very efficient since the set

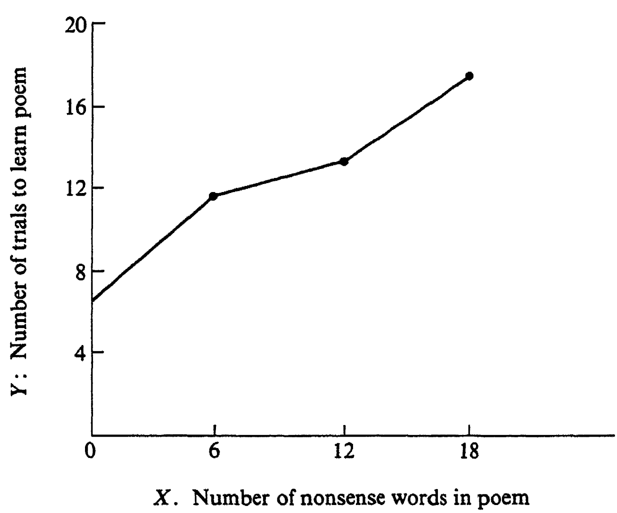
of alternative hypotheses can be restricted to specify that the expected values satisfy a linear equation with positive slope. While it will not be shown, it is reasonable to assume that the omnibus F test is less powerful than a test specifically designed to test for the existence of a linear relationship. For that reason, classical regression theory is more powerful. In Section 20-10, it is presented within the framework of the analysis of variance.

20-3 ESTIMATION OF THE REGRESSION LINE: METHOD OF LEAST SQUARES

Consider a general regression and let X represent the independent variable and Y represent the dependent variable. On each of the N individuals included in the sample let the following ordered pairs of values (x_i, y_i), $i = 1, 2, \ldots, N$, be observed. For the observed outcomes of the study of Section 20-2, the 32 outcomes are shown in Figure 20-5, the scatter diagram of the study.

Let it be assumed that the expected regression line is given by $Y_X = \alpha + \beta X$. Note that this equation contains two unknown parameters: α, the Y intercept of the equation, and β, the slope of the equation. From elementary algebra it is known that if these two unknown quantities could be related to one another by two simultaneous linear equations, they could be estimated with ease. Fortunately, it is easy to obtain two equations that involve these two unknown parameter values and all of the observed values of X and Y. To obtain the first of these equations, take each individual pair of observations and substitute them directly into the equation and

Figure 20-4. Empirical regression line of Y on X for the data of Table 20-1.



add the resulting set of equations. Thus, for the first equation and the N subjects, the following set of equations generates Equation 1.

Subject	Equation Set 1
1	$y_1 = \alpha + \beta x_1$
2	$y_2 = \alpha + \beta x_2$
$\cdot$	$\cdot$
i	$y_i = \alpha + \beta x_i$
	$\cdot$
N	$y_N = \alpha + \beta x_N$

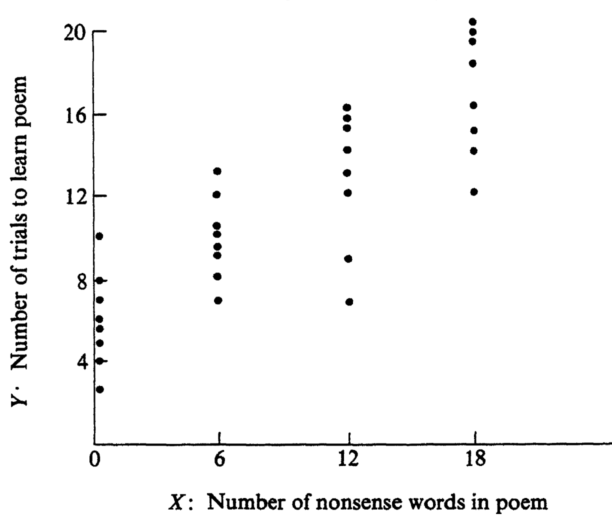
Equation 1: $\sum_{i=1}^{N} y_i = N\alpha + \beta \sum_{i=1}^{N} x_i$

To obtain the second equation, take each of the equations of the above set, multiply it by its corresponding x_i, and then add the resulting set of equations This gives the following set of equations, which generate Equation 2.

Subject	Equation Set 2
1	$y_1 x_1 = \alpha x_1 + \beta x_1^2$
2	$y_2 x_2 = \alpha x_2 + \beta x_2^2$
$\vdots$	$\vdots$
i	$y_i x_i = \alpha x_i + \beta x_i^2$
$\vdots$	$\vdots$
N	$y_N x_N = \alpha x_N + \beta x_N^2$

Equation 2: $\sum_{i=1}^{N} y_i x_i = \alpha \sum_{i=1}^{N} x_i + \beta \sum_{i=1}^{N} x_i^2$

Figure 20-5. The scatter diagram of the study



Thus, two equations have been generated that can be solved for α and β. From Equation 1, it follows that

$$\begin{aligned}\alpha &= \frac{1}{N}\left[\sum_{i=1}^{N} y_i - \beta \sum_{i=1}^{N} x_i\right] \\ &= \frac{1}{N}\sum_{i=1}^{N} y_i - \beta\frac{1}{N}\sum_{i=1}^{N} x_i \\ &= \bar{Y} - \beta\bar{X}\end{aligned}$$

Substituting this result into Equation 2, we find that

$$\sum_{i=1}^{N} y_i x_i = (\bar{Y} - \beta\bar{X}) \sum_{i=1}^{N} x_i + \beta \sum_{i=1}^{N} x_i^2$$

Solving this equation for β, one finds that

$$\beta\left[\sum_{i=1}^{N} x_i^2 - \bar{X}\sum_{i=1}^{N} x_i\right] = \left[\sum_{i=1}^{N} y_i x_i - \bar{Y}\sum_{i=1}^{N} x_i\right]$$

so that

$$\beta = \frac{\sum_{i=1}^{N} y_i x_i - \bar{Y}\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2 - \bar{X}\sum_{i=1}^{N} x_i} = \frac{N\left(\sum_{i=1}^{N} y_i x_i\right) - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)}{N\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i\right)^2}$$

It is customary to denote the estimated values by $\hat{\alpha}$ and $\hat{\beta}$. These last two results show that the estimate of the regression line is given by

$$\hat{Y}_X = \hat{\alpha} + \hat{\beta}X = (\bar{Y} - \hat{\beta}\bar{X}) + \hat{\beta}X = \bar{Y} + \hat{\beta}(X - \bar{X})$$

In this form, it is seen that it is necessary to compute only the average value of the X and Y variables and the estimate of the slope. The estimation of α can be ignored.

The quantity $N(\sum_{i=1}^{N} x_i^2) - (\sum_{i=1}^{N} x_i)^2$ appears frequently in the formulas of linear regression theory and so it is convenient to have a special notation for it. In this text, Δ_X will be used to denote this algebraic quantity. Note that Δ_X is based upon known *constant* values of X and is therefore not a random variable. The only random variable of the study is Y. With this notation,

$$\hat{\beta} = \frac{1}{\Delta_X}\left[N\left(\sum_{i=1}^{N} y_i x_i\right) - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)\right]$$

These estimates of α and β generated by the two artificially constructed equations can be derived from a "classical theory" of estimation called least squares. The estimates themselves are sometimes called least-square estimators. For this procedure, one considers the squared deviation of each y_i from the regression line or its predicted value. Each of these squared deviations is added, and then α and β are chosen so as to make the sum of the squared deviations minimum. The determination

of these estimates is a simple problem in differential calculus. Since little is gained in deriving these estimates under this model, the following theorem is stated without proof.

Theorem 20-1

The least-square estimates of α and β for the linear regression line $Y_X = \alpha + \beta X$ based on a random sample of N ordered pairs (x_i, y_i), $i = 1, 2, \ldots, N$, are given by

$$\hat{\alpha} = \frac{\left(\sum_{i=1}^{N} y\right)\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i y_i\right)\left(\sum_{i=1}^{N} x_i\right)}{N\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i\right)^2}$$

$$\hat{\beta} = \frac{N\left(\sum_{i=1}^{N} x_i y_i\right) - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)}{N\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i\right)^2}$$

These results can now be used to estimate α and β for the data of Table 20-1. By definition,

$$\begin{aligned}\sum_{i=1}^{N} x_i &= (0+0+0+0+0+0+0+0)\\ &\quad + (6+6+6+6+6+6+6+6)\\ &\quad + (12+12+12+12+12+12+12+12)\\ &\quad + (18+18+18+18+18+18+18+18)\\ &= 288\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{N} y_i &= (4+8+7+6+10+6+3+5)\\ &\quad + (9+8+12+10+7+13+9+10)\\ &\quad + (7+15+12+9+16+14+13+15)\\ &\quad + (15+18+20+14+12+19+19+16)\\ &= 49+78+101+133 = 361\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{N} x^2 &= (0^2+0^2+\cdots+0^2)+\cdots+(18^2+18^2+\cdots+18^2)\\ &= 4{,}032\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{N} y^2 &= (4^2+8^2+\cdots+5^2)+\cdots+(15^2+18^2+\cdots+16^2)\\ &= 4{,}735\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{N} x_i y_i &= [0(4)+0(8)+\cdots+0(5)]+\cdots+[18(15)+18(18)+\cdots+18(16)]\\ &= 4{,}074\end{aligned}$$

For this set of data,

$$\hat{\beta} = \frac{32(4{,}074) - 288(361)}{32(4{,}032) - 288(288)} = \frac{26{,}400}{46{,}080} = .5729$$

$$\bar{X} = \tfrac{288}{32} = 9$$

$$\bar{Y} = \tfrac{361}{32} = 11.2813$$

$$\hat{\alpha} = 11.2813 - .5729(9) = 6.1252$$

and

$$\hat{Y}_X = 6.152 + .5729X$$

Thus, it takes approximately six trials to learn the poem with no nonsense words. For each nonsense word added to the poem, the number of trials to learn increases by about one-half trial.

For $X = 18$, $\hat{Y}_{18} = 6.1252 + .5729(18) = 16.4374$
For $X = 12$, $\hat{Y}_{12} = 6.1252 + .5729(12) = 13\ 0000$
For $X = 6$, $\hat{Y}_{6} = 6.1252 + .5729(6) = 9.5626$
For $X = 0$, $\hat{Y}_{0} = 6.1252 + .5729(0) = 6.1252$

These estimates are quite close to the sample averages. As will be seen later, these estimates are efficient and also unbiased.

20-4 PROPERTIES OF THE LEAST-SQUARE ESTIMATORS

The basic properties of the least-square estimates are summarized in Theorems 20-2 through 20-6, whose proofs may be skipped over by the uninterested reader.

Theorem 20-2

$\hat{\beta}$, the least-square estimator of the slope, is an unbiased estimate of β

Proof. As has been shown in Section 20-3,

$$\hat{\beta} = \frac{1}{\Delta_X}\left[N\left(\sum_{i=1}^{N} x_i y_i\right) - N\bar{X}\left(\sum_{i=1}^{N} y_i\right)\right]$$

$$= \frac{N}{\Delta_X}[(x_1 y_1 + x_2 y_2 + \cdots + x_N y_N) - \bar{X}(y_1 + y_2 + \cdots + y_N)]$$

$$= \frac{N}{\Delta_X}[(x_1 - \bar{X})y_1 + (x_2 - \bar{X})y_2 + \cdots + (x_N - \bar{X})y_N]$$

According to Theorem 6-1, the expected value of a sum of random variables is the sum of the expected values. Furthermore, since $E(y_i) = Y_{x_i} = \alpha + \beta x_i$, it follows that

$$\begin{aligned}
E(\hat{\beta}) &= \frac{N}{\Delta_X}[(x_1 - \bar{X})E(y_1) + (x_2 - \bar{X})E(y_2) + \cdots + (x_N - \bar{X})E(y_N)] \\
&= \frac{N}{\Delta_X}[(x_1 - \bar{X})(\alpha + \beta x_1) + (x_2 - \bar{X})(\alpha + \beta x_2) + \cdots + (x_N - \bar{X})(\alpha + \beta x_N)] \\
&= \frac{N}{\Delta_X}\left[\sum_{i=1}^{N}(x_i - \bar{X})\alpha + \sum_{i=1}^{N}(x_i - \bar{X})(x_i\beta)\right] \\
&= \frac{N}{\Delta_X}\left[\alpha\sum_{i=1}^{N}(x_i - \bar{X}) + \beta\sum_{i=1}^{N}(x_i^2 - \bar{X}x_i)\right] \\
&= \frac{N}{\Delta_X}\left[\alpha(0) + \beta\sum_{i=1}^{N}(x_i^2 - x_i\bar{X})\right] \\
&= \frac{N}{\Delta_X}\beta\left[\sum_{i=1}^{N}\left\{x_i^2 - \frac{\left(\sum_{i=1}^{N}x_i\right)}{N}x_i\right\}\right] \\
&= \frac{N}{\Delta_X}\beta\left[\frac{N\sum_{i=1}^{N}x_i^2 - \left(\sum_{i=1}^{N}x_i\right)\left(\sum_{i=1}^{N}x_i\right)}{N}\right] \\
&= \beta\left[\frac{1}{\Delta_X}\left\{N\sum_{i=1}^{N}x_i^2 - \left(\sum_{i=1}^{N}x_i\right)^2\right\}\right] \\
&= \beta\frac{1}{\Delta_X}\Delta_X = \beta
\end{aligned}$$

This completes the proof.

Theorem 20-3

The least-square estimator of the Y intercept $\hat{\alpha}$ is an unbiased estimator of α.

Proof. By derivation,

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

According to Theorems 6-1 and 7-3,

$$\begin{aligned}
E(\hat{\alpha}) &= E(\bar{Y}) - E(\hat{\beta}\bar{X}) \\
&= (\alpha + \beta\bar{X}) - \bar{X}E(\hat{\beta}) \\
&= \alpha + \beta\bar{X} - \beta\bar{X} \\
&= \alpha
\end{aligned}$$

This completes the proof.

Theorem 20-4

The least-square estimate of the regression line $\hat{Y}_X$ is an unbiased estimator of Y_X.

Proof. By derivation,

$$\hat{Y}_X = \hat{\alpha} + \hat{\beta}X$$

According to Theorems 6-1, 7-3, 20-2, and 20-3,

$$\begin{aligned} E(\hat{Y}_X) &= E(\hat{\alpha}) + E(\hat{\beta}X) \\ &= \alpha + \beta X \\ &= Y_X \end{aligned}$$

This completes the proof.

20-5 THE VARIANCES OF THE SAMPLING DISTRIBUTION OF $\hat{\beta}$ AND $\hat{Y}_X$

Up to now, the exposition of regression theory has depended only upon the assumptions of linearity of regression and the selection of independent samples. To continue the exposition, it now becomes necessary to add the analysis-of-variance assumption of common variance for each of the individual distributions. With the assumption that $\mathrm{Var}(Y_X) = \sigma^2_{Y.X}$ is constant for all values of X, it is possible to determine the variance of the distribution of $\hat{\beta}$.

Theorem 20-5

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2_{Y.X}}{\sum_{i=1}^{N} (x_i - \bar{X})^2} = \frac{N}{\Delta_X}\sigma^2_{Y.X}$$

Proof. By derivation,

$$\hat{\beta} = \frac{N}{\Delta_X}[(x_1 - \bar{X})y_1 + (x_2 - \bar{X})y_2 + \cdots + (x_N - \bar{X})y_N]$$

Since the y_i are statistically independent, it follows from Theorem 6-4 that the variance of a sum of random variables is the sum of the variances and therefore

$$\begin{aligned} \mathrm{Var}(\hat{\beta}) &= \frac{N^2}{\Delta_X^2}[(x_1 - \bar{X})^2\mathrm{Var}(y_1) + (x_2 - \bar{X})^2\,\mathrm{Var}(y_2) + \cdots + (x_N - \bar{X})^2\,\mathrm{Var}(y_N)] \\ &= \frac{N^2}{\Delta_X^2}\sigma^2_{Y.X}[(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + \cdots + (x_N - \bar{X})^2] \\ &= \frac{N^2}{\Delta_X^2}\sigma^2_{Y.X}\sum_{i=1}^{N}(x_i - \bar{X})^2 \\ &= \frac{N^2}{\Delta_X^2}\sigma^2_{Y.X}\left[\frac{N\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i\right)^2}{N}\right] \\ &= \frac{N^2}{\Delta_X^2}\sigma^2_{Y.X}\frac{\Delta_X}{N} \\ &= \frac{N}{\Delta_X}\sigma^2_{Y.X} \end{aligned}$$

This completes the proof.

:e the discussion continues to use the regression equation in the form $\bar{Y} + \hat{\beta}(X - \bar{X})$, it won't be necessary to determine the Var($\hat{\alpha}$). Instead, all ıe needs to know is that

$$(\bar{Y}) = \frac{\sigma^2_{Y.X}}{N}$$

'er, for completeness, the following is stated without proof.

20-6

$$\hat{\alpha}) = \frac{\sigma^2_{Y.X}}{\sum_{i=1}^{N}(x_i - \bar{X})^2}\left[\frac{\sum_{i=1}^{N} x_i^2}{N}\right] = \frac{\sum_{i=1}^{N} x_i^2}{N}\,\mathrm{Var}(\hat{\beta})$$

lly, the basic result follows in Theorem 20-7.

ı 20-7

$$\hat{Y}_X) = \frac{\sigma^2_{Y.X}}{N}\left[1 + \frac{N^2(X - \bar{X})^2}{\Delta_X}\right]$$

By definition,

$$\hat{Y}_X) = \mathrm{Var}[\bar{Y} + \hat{\beta}(X - \bar{X})]$$

ırems 6-4 and 7-4,

$$= \mathrm{Var}(\bar{Y}) + (X - \bar{X})^2\,\mathrm{Var}(\hat{\beta})$$

$$= \frac{\sigma^2_{Y.X}}{N} + (X - \bar{X})^2\,\sigma^2_{Y.X}\frac{N}{\Delta_X}$$

$$= \frac{\sigma^2_{Y.X}}{N}\left[1 + \frac{N^2(X - \bar{X})^2}{\Delta_X}\right]$$

ıpletes the proof.

:riving this last formula it has been implicitly assumed that the sampling tions of $\bar{Y}$ and $\hat{\beta}$ are statistically independent. While this is, indeed, true, no is made to prove it.

neral, $\sigma^2_{Y.X}$ is unknown so that all of these variances cannot be determined. ıs is basically an analysis-of-variance design it might seem that the pooled or stimate of the common variance could be used as an estimate of $\sigma^2_{Y.X}$. While :ceptable, it is not recommended since there is an estimator of $\sigma^2_{Y.X}$ that has efficiency. This estimator is easy to derive.

two parameters are estimated from the data instead of $(K - 1)$, as for the analysis of variance, it is known that

$$\frac{1}{N-2}\sum_{i=1}^{N}(y_i - \hat{Y}_{x_i})^2$$

should be an estimate of $\sigma^2_{\hat{Y}.X}$ based on $(N-2)$ degrees of freedom. With considerable algebra, one can show that this estimate is given by the formula of Theorem 20-8, which is stated without proof but which is given as an exercise at the end of this chapter.

Theorem 20-8

An unbiased estimate of $\sigma^2_{\hat{Y}.X}$ is given by

$$S^2_{\hat{Y}.X} = \frac{N-1}{N-2}\,[S^2_{\hat{Y}} - \hat{\beta}^2 S^2_{\hat{X}}]$$

For the data of this study,

$$S^2_{\hat{Y}} = \frac{32(4{,}735) - (361)^2}{32(31)} = 21.3700$$

$$S^2_{\hat{X}} = \frac{32(4{,}032) - (288)^2}{32(31)} = 46.4516$$

$$S^2_{\hat{Y}.X} = \tfrac{31}{30}[21.3700 - (.5729)^2\,(46.4516)] = 6.3282$$

For the data of Table 20-1, MSW is equal to 6.7188. Even though the difference is very small, $S^2_{\hat{Y}.X}$ is still the preferred estimate.

For the data of this study,

$$\mathrm{SE}^2_{\hat{\beta}} = \frac{N}{\Delta_X} S^2_{\hat{Y}.X} = \frac{32(6.3282)}{46{,}080} = .00439$$

$$\mathrm{SE}^2_{\hat{\alpha}} = \frac{\sum_{i=1}^{N} x_i^2}{N}\,\mathrm{SE}^2_{\hat{\beta}} = \frac{4{,}032}{32}(.00439) = .5537$$

$$\mathrm{SE}^2_{\hat{Y}} = \frac{S^2_{\hat{Y}.X}}{N} = \frac{6.3282}{32} = .19775$$

$$\mathrm{SE}^2_{\hat{Y}_X} = \frac{S^2_{\hat{Y}.X}}{N}\left[1 + \frac{N^2(X-\bar{X})^2}{\Delta_X}\right] = \frac{6.3282}{32}\left[1 + \frac{(32)^2\,(X-9)^2}{46{,}080}\right]$$

$$\mathrm{SE}^2_{\hat{Y}_0} = \frac{6.3282}{32}\left[1 + \frac{(32)^2\,(0-9)^2}{46{,}080}\right] = .5537$$

$$\mathrm{SE}^2_{\hat{Y}_6} = \frac{6.3282}{32}\left[1 + \frac{(32)^2\,(6-9)^2}{46{,}080}\right] = .2373$$

$$\mathrm{SE}^2_{\hat{Y}_{12}} = \frac{6.3282}{32}\left[1 + \frac{(32)^2\,(12-9)^2}{46{,}080}\right] = .2373$$

$$\mathrm{SE}^2_{\hat{Y}_{18}} = \frac{6.3282}{32}\left[1 + \frac{(32)^2\,(18-9)^2}{46{,}080}\right] = .5537$$

For the individual group means,

$$\mathrm{SE}^2_{\bar{Y}} = \frac{S^2_{Y \cdot X}}{N_0} = \frac{6.3282}{8} = .7910$$

As can be seen, the standard errors based on the regression equation are considerably smaller than are the estimates based on the individual group means. Clearly, $.5537 < .7910$ and $.2373 < .7910$

20-6 CONFIDENCE INTERVALS FOR THE UNKNOWN PARAMETERS

The confidence intervals for the parameters of a regression model are given by

$$\hat{\beta} - t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\beta}} < \beta < \hat{\beta} + t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\beta}}$$

$$\hat{\alpha} - t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\alpha}} < \alpha < \hat{\alpha} + t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\alpha}}$$

$$\hat{Y}_X - t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{Y}_X} < Y_X < \hat{Y}_X + t_\nu\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{Y}_X}$$

$$\frac{(N-2)\,S^2_{Y \cdot X}}{\chi^2_\nu(1-\alpha/2)} < \sigma^2_{Y \cdot X} < \frac{(N-2)\,S^2_{Y \cdot X}}{\chi^2_\nu(\alpha/2)}$$

where $\nu = N - 2$.

For the data of the study, $\nu = N - 2 = 32 - 2 = 30$, and $t_{30}(.025) = -2.042$. Thus, the 95 percent confidence intervals are given by

$$.5729 - 2.042\sqrt{.00439} < \beta < .5729 + 2.042\sqrt{.00439}$$

$$.4381 < \beta < .7077$$

$$6.1252 - 2.042\sqrt{.5537} < \alpha < 6.1252 + 2.042\sqrt{.5537}$$

$$4.6657 < \alpha < 7.6447$$

$$6.1252 - 2.042\sqrt{.5537} < Y_0 < 6.1252 + 2.042\sqrt{.5537}$$

$$4.6657 < Y_0 < 7.6447$$

$$9.5626 - 2.042\sqrt{.2373} < Y_6 < 9.5626 + 2.042\sqrt{.2373}$$

$$8.5677 < Y_6 < 10.5575$$

$$13.0000 - 2.042\sqrt{2373} < Y_{12} < 13.0000 + 2.042\sqrt{.2373}$$

$$12.0051 < Y_{12} < 13.9949$$

$$16.4373 - 2.042\sqrt{.5537} < Y_{18} < 16.4373 + 2.042\sqrt{.5537}$$

$$14.9178 < Y_{18} < 17.9569$$

$$\frac{(30)(6.3282)}{46.98} < \sigma^2_{Y.X} < \frac{(30)(6.3282)}{16.79}$$

$$4.0409 < \sigma^2_{Y.X} < 11.3071$$

$$2.02 < \sigma_{Y.X} < 3.36$$

The confidence bands for the regression line are shown in Figure 20-6. There is reason to believe that the unknown regression line is somewhere within the drawn band with a probability of a type I error controlled at $\alpha = .05$.

20-7 TESTS OF HYPOTHESIS FOR THE REGRESSION MODEL

The tests of hypothesis and test statistics for the regression model are based on the usual t and χ^2 models as follows:

1. To test $H_0: \beta = \beta_0$ versus $H_1: \beta \neq \beta_0$, the appropriate test statistic is given by

$$t = \frac{(\hat{\beta} - \beta_0)}{\mathrm{SE}_{\hat{\beta}}}$$

H_0 is rejected if $t < t_\nu(\alpha/2)$ or if $t > t_\nu(\alpha/2)$, where $\nu = N - 2$.

2. To test $H_0: \alpha = \alpha_0$ versus $H_1: \alpha \neq \alpha_0$, the appropriate test statistic is given by

$$t = \frac{\hat{\alpha} - \alpha_0}{\mathrm{SE}_{\hat{\alpha}}}$$

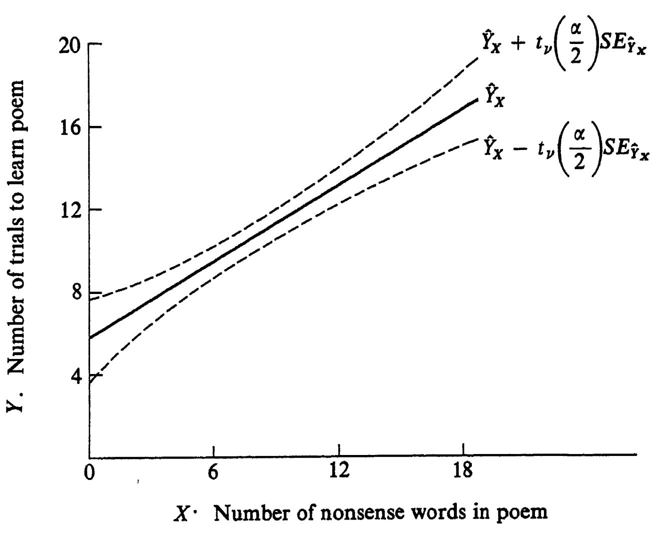
H_0 is rejected if $t < t_\nu(\alpha/2)$ or if $t > t_\nu(\alpha/2)$, where $\nu = N - 2$.

3. To test $H_0: \sigma^2_{\hat{Y}.X} = \sigma^2_0$ versus $H_1: \sigma^2_{\hat{Y}\,X} \neq \sigma^2_0$, the appropriate test statistic is given by

$$\chi^2 = \frac{(N-2)\, S^2_{\hat{Y}.X}}{\sigma^2_{\hat{Y}\,X}}$$

H_0 is rejected if $\chi^2 < \chi^2_\nu(\alpha/2)$ or if $\chi^2 > \chi^2_\nu(1 - \alpha/2)$, where $\nu = N - 2$.

Figure 20-6. Confidence bands for the regression line $Y = \alpha + \beta X$ for a type I error of .05.



The assumptions for these tests and for the confidence interval are the usual analysis-of-variance assumptions:

1. Random samples on each of the K populations
2. Common variance for each population criterion variable
3. Normality of the Y variable or the N_k sufficiently large so that the conditional distribution of the $\bar{Y}_x$ are approximately normal

In addition, there are two new assumptions that must be included. These assumptions are:

4. The X values are constants that are assigned before the data is collected.
5. The expected values lie on a straight line or are algebraically defined by a linear equation.

To illustrate the use of these tests, suppose it were hypothesized that $E(Y_X) = 10 + 2X$ and $\sigma^2_{Y.X} = 4$. The resulting tests are as follows:

1. For testing $H_0: \beta = 2$ versus $H_1: \beta \neq 2$, the value of the test statistic is given by

$$t = \frac{.5729 - 2}{.066} = -21.8$$

   H_0 is rejected since $t_{30}(.025) = -2.045$ and $t_{30}(.975) = 2.045$.
2. For testing $H_0: \alpha = 10$ versus $H_1: \alpha \neq 10$, the value of the test statistic is given by

$$t = \frac{6.1252 - 10}{.7441} = -5.21$$

   H_0 is rejected since $t_{30}(.025) = -2.045$ and $t_{30}(.975) = 2.045$.
3. For testing $H_0: \sigma^2_{Y.X} = 4$ versus $H_1: \sigma^2_{Y.X} \neq 4$, the value of the test statistic is given by

$$\chi^2 = \frac{30(6.3282)}{4} = 47.46$$

   H_0 is rejected since $\chi^2_{30}(.025) = 14.95$ and $\chi^2_{30}(.975) = 46.98$.

20-8 REGRESSION THEORY WHEN THE X VARIABLE IS NOT SPECIFIED IN ADVANCE

A frequent occurrence in behavioral research is that the X variable may not be specified in advance of the experiment. Frequently, the X variable is measured or

observed simultaneously with the Y variable. As an example of this situation, consider the following example.

Eighteen Protestant clergymen were given a test designed to measure their degree of philosophical and religious independence, with low scores indicating low levels of independence. They were also given a masculinity-femininity test, with low scores correlated with feminine characteristics. The question to be asked was: "Is there a relationship between philosophical independence and masculinity-femininity as measured by these two tests?" In this case, the X variable is the masculinity-femininity score and the Y variable is the independence score. While it is true that the X variable is truly a random variable, one can assume that the X variables represent a parametric value for each individual and are thereby without error While this may be a rather strong assumption, the results that would be obtained by selecting a different random sample will in general be very close to the results that were obtained for the actual random sample selected. The scores for the 18 clergymen are shown in

Table 20-2. Scores on the X and Y variables for the 18 clergymen.

X *M–F score*	Y: *Autonomy score*
8	12
8	22
9	30
10	31
10	40
13	18
15	22
20	27
22	20
23	19
23	26
23	40
25	28
26	42
30	30
32	51
40	40
52	37

Table 20-2. The scatter diagram for the data is shown in Figure 20-7. In this case it appears as though the linear relationship, if it exists, is very weak. For these data,

1. $N = 18$

2. $\sum_{i=1}^{18} x_i = 389$

3. $\sum_{i=1}^{18} y_i = 535$

4. $\sum_{i=1}^{18} x_i y_i = 12{,}599$

5. $\sum_{i=1}^{18} x_i^2 = 10{,}803$

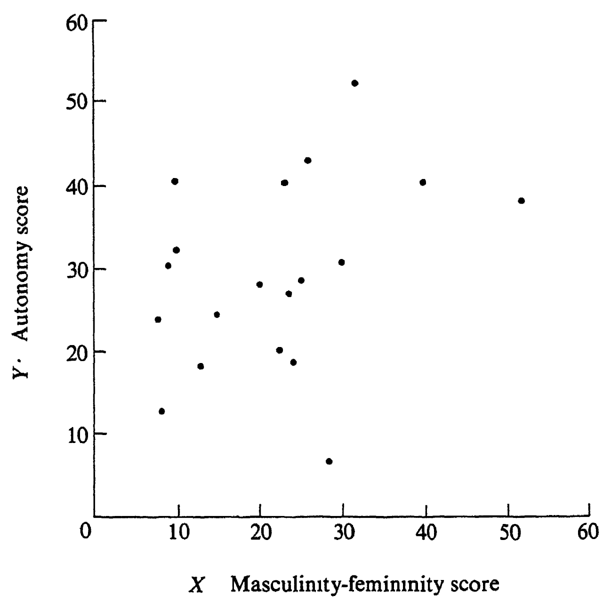
6. $\sum_{i=1}^{18} y_i^2 = 17{,}681$

7. $\bar{X} = \frac{389}{18} = 21.6111$

8. $\bar{Y} = \frac{535}{18} = 29.7222$

9. $\hat{\beta} = \dfrac{18(12{,}599) - (389)(535)}{18(10{,}803) - (389)(389)} = 4328$

Figure 20-7. Scatter diagram for the 18 clergymen of Table 20-2.



10. $\hat{Y}_X = \bar{Y} + \hat{\beta}(X - \bar{X}) = 29.7222 + 4328(X - 21.6111)$
$$= 20.3689 + .4328(X)$$

11. $S_Y^2 = \dfrac{18(17{,}681) - (535)^2}{18(17)} = 104.6830$

12. $S_X^2 = \dfrac{18(10{,}803) - (389)^2}{18(17)} = 140.9575$

13. $S_{Y\,X}^2 = \dfrac{N-1}{N-2}[S_Y^2 - \hat{\beta}^2 S_X^2]$
$$= \tfrac{17}{16}[104.6830 - (.4328)^2(140.9575)]$$
$$= 83.1713$$

If there is no linear relationship between the two variables, one would expect β to equal 0. Thus, the test of no linear relationship is given by

$$t = \frac{\hat{\beta} - 0}{\mathrm{SE}_\beta}$$

For these data,

$$\mathrm{SE}_\beta^2 = \frac{N}{\Delta_X} S_{Y.X}^2 = \frac{18}{43{,}133}(83.1713) = .0347$$

and

$$t = \frac{.4328}{\sqrt{.0347}} = \frac{.4328}{.1863} = 2.32$$

Since $t = 2.32 > t_{16}(.975) = 2.2$, H_0 is rejected. There is reason to believe that the two variables are related. The best linear estimate of the relationship is given by

$$\hat{Y}_X = 20.3689 + .4328(X)$$

20-9 EXPLAINED VARIANCE FOR REGRESSION MODELS

For regression models, it was seen that the unbiased estimate of the residual or unexplained variance is given by

$$S_{Y.X}^2 = \frac{N-1}{N-2}[S_Y^2 - \hat{\beta}^2 S_X^2]$$

This suggests that

$$(N-2)\,S_{Y.X}^2 = (N-1)\,S_Y^2 - (N-1)\hat{\beta}^2 S_X^2$$

so that

$$(N-1)\,S_Y^2 = (N-2)\,S_{Y.X}^2 + (N-1)\hat{\beta}^2 S_X^2$$

or

$$\sum_{i=1}^{N}(y_i - \bar{Y})^2 = \sum_{i=1}^{N}(y_i - \hat{Y}_{X_i})^2 + \sum_{i=1}^{N}(\hat{Y}_{X_i} - \bar{Y})^2$$

or that

$$\text{SST} = \text{SSU} + \text{SSE}$$

Thus, a measure of explained variance is given by

$$\hat{\omega}^2 = \frac{\text{SSE}}{\text{SST}} = \frac{(N-1)\,\hat{\beta}^2 S_X^2}{(N-1)\,S_Y^2} = \hat{\beta}^2 \frac{S_X^2}{S_Y^2}$$

When $\hat{\beta} = 0$, the regression is parallel to the X axis and all conditional distributions are identical, indicating that corresponding conditional probabilities are identical.

For the two regression examples, the estimates of ω^2 are easy to obtain. For the first example,

$$\hat{\omega}^2 = \frac{(.5729)^2(46.4516)}{21.3700} = .7134$$

indicating a very strong linear relationship between the number of nonsense words in the poem and the number of trials it takes to learn the poem. For the second example,

$$\hat{\omega}^2 = \frac{(.4328)^2(140.9575)}{104.6830} = .2522$$

indicating a moderate linear relationship between the masculinity-femininity and autonomy scales.

20-10 REGRESSION AS AN ANALYSIS-OF-VARIANCE MODEL

The basic elements of a regression model treated as an analysis-of-variance model have already been presented. In the one-way analysis-of-variance model, one begins with

$$\text{SST} = \text{SSB} + \text{SSW}$$

and

$$(N-1) = (K-1) + (N-K)$$

and from these algebraic quantities one sets up an analysis-of-variance table.

A little reflection will show that the $Y_X = \alpha + \beta X$ for each value of X correspond to the μ_k of an analysis-of-variance model. In fact, $\mu_k = \alpha + \beta X_k$. Thus, SSB measures the part of the variability that exists between the Y_X. If the relationship between X and Y is entirely linear, SSB measures it completely. However, if the regression is more than linear, SSB also measures this extra component. This suggests that SSB measures two components, and such is the case. If SS($\hat{\beta}$) denotes the sum of

squares associated with the linear component and if $SS(D\hat{\beta})$ denotes the sum of squares associated with the curvilinear component or the deviation from linearity, one can show that

$$SSB = SS(\hat{\beta}) + SS(D\hat{\beta})$$

with

$$(K-1) = 1 + (K-2)$$

With a little effort one can prove the following theorem, which is presented without proof.

Theorem 20-9

The sum of squares associated with the linear component of a regression model is given by

$$SS(\hat{\beta}) = (N-1)\hat{\beta}^2 S_X^2$$

The proof of this result is given at the end of this chapter as an exercise. Thus, for the data of Table 20-1,

$$SS(\hat{\beta}) = 31(.5729)^2(46.4516) = 472.6229$$

For these same data,

$$\text{I} = \sum_{k=1}^{4} \sum_{i=1}^{8} y_{ik}^2 = 4{,}735$$

$$\text{II} = \frac{1}{N} T_{..}^2 = \tfrac{1}{32}(361)^2 = 4{,}072\ 5313$$

$$\text{III}_C = \sum_{i=1}^{4} \frac{T_{.k}^2}{N_0} = \tfrac{1}{8}(49^2 + 78^2 + 101^2 + 133^2) = 4{,}546.8750$$

$$SSB = \text{III}_C - \text{II} = 474.3437$$

$$SS(D\hat{\beta}) = SSB - SS(\hat{\beta}) = 474.3437 - 472.6229 = 1.7208$$

$$SSW = \text{I} - \text{III}_C = 4{,}735 - 4{,}546.8750 = 188.1250$$

$$SST = \text{I} - \text{II} = 4{,}735 - 4{,}072.5313 = 662.4687$$

From these figures, the analysis-of-variance table is given as shown in Table 20-3.

Since $SSB = SS(\hat{\beta}) + SS(D\hat{\beta})$, it is customary to list them as subheadings under the "between groups" source of variance. In addition, the degree of freedom for each of the individual components is also shown. Furthermore, since no interest exists in testing $H_0: \mu_1 = \mu_2 = \cdots = \mu_K = \mu_0$, the F ratio for this test is ignored.

Table 20-3. The analysis-of-variance table for the data of Table 20-1.

Source of variation	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F ratio*
Between groups	3	474.3437		
Linear	1	472 6229	472 6229	70 34
Deviation from linearity	2	1.7208	.8604	13
Within groups	28	188 1250	6 7188	
Total	31	662.4687		

On the other hand, the test statistic for testing H_0: no relationship between X and Y versus H_1: linear relationship between X and Y is given by

$$F = \frac{\text{MS}(\hat{\beta})}{\text{MSW}}$$

The hypothesis of no relationship is rejected if $F > F_{1,N-K}(1-\alpha) = F_{1,28}(.95) = 4\ 20$. Since $F = 70.34 > 4.20$, the hypothesis of no relationship is rejected and it is concluded that a linear component exists to the relationship between X and Y.

The test statistic for testing H_{0C}: no curvilinear component to the relationship between X and Y versus H_{1C}: curvilinear component exists is given by

$$F = \frac{\text{MS}(D\hat{\beta})}{\text{MSW}}$$

The hypothesis of no curvilinear component is rejected if $F > F_{K-2,N-K}(1-\alpha) = F_{2,28}(.95) = 3.34$. Since $F = .13 < 3.34$, the hypothesis is not rejected. Thus, it is concluded that X and Y are linearly related.

It should be noted that regression analysis performed as an analysis of variance does not require that K independent groups be established prior to data collection. However, the degrees of freedom and sum of squares do not partition in exactly the same manner as they do for the analysis-of-variance model. In addition, since the sum of squares within groups does not exist, it is customary to call the corresponding sum of squares the sum of squares residual, and the estimate of $\sigma^2_{Y.X}$ is called the mean square residual. The appropriate partitioning for the example of Section 20-8 is as follows:

1. $\text{I} = 17{,}681$
2. $\text{II} = \frac{1}{18}(535)^2 = 15{,}901.39$
3. $\text{SST} = \text{I} - \text{II} = 17{,}681 - 15{,}901.39 = 1{,}779.61$
4. $\text{SSR} = (N-2)\,S^2_{Y.X} = 16(83\ 1713) = 1{,}330.74$
5. $\text{SS}(\hat{\beta}) = (N-1)\,\hat{\beta}^2 S^2_X = 17(.4328)^2(140.9575)$
   $= 448.86$

Table 20-4. Analysis-of-variance table for $K=1$ in a regression model.

Source of variance	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F ratio*
Linear component	1	448.86	448.86	5 40
Residual	16	1,330.74	83 17	
Total	17	1,779.60		

so that the analysis-of-variance table is given by Table 20-4. In this case, a test for a curvilinear component is not available since there are no degrees of freedom between groups that can be partitioned. Since $F=5.40>F_{1,16}(.95)=4.49$, the hypothesis of no relationship is rejected. It should be noted that this analysis is identical to the analysis presented in Section 20-8. For the test of H_0: $\beta=0$, it was seen that $t=2.32$. Since $t_\nu^2=F_{1,\nu}$, it is to be expected that $t^2=(2.32)^2=5.40$ $=F$. Except for rounding-off errors, the condition does hold.

20-11 THE K-SAMPLE REGRESSION MODEL (THE ANALYSIS OF COVARIANCE)

Regression models are commonly employed in comparative research studies As an example of a two-sample regression model, consider a study that involves a comparison of a standard diet versus a diet rich in proteins, fed over a period of six weeks to new-born maze-bright rats. If the dependent variable of the study is gain in weight over the six-week period, one would have to consider and control the effects that initial weight has upon the weight gain. For example, it might be that weight gain is positively correlated with initial weight, and as a consequence, inequalities in rat birth weight might confound the results and interpretation of the study. One way to statistically control for differences in initial weights is to match the rats by weight before random assignment to the standard or to the new diet. When matching is physically impossible, one may use regression theory to achieve statistical control by what is called an analysis-of-covariance design. This procedure is illustrated in the following hypothetical study on new methods of reading-instruction.

Three fourth-grade classrooms in a middle-class suburban school were assigned one of three different reading methods. Students in classroom 1 were taught according to a traditional reading program. Students in classroom 2 were trained under the traditional method with extra emphasis on phonic methods of instruction. Students in classroom 3 utilized the traditional program along with special training in the "look-say" method of reading instruction. At the end of a six-week period, students were given a reading comprehension test. The test scores, together with the

Table 20-5. Observed data and sample statistics for the three classrooms of the fourth-grade reading study.

Characteristic teaching method	*Sample 1* *Standard*		*Sample 2* *Phonics*		*Sample 3* *Look-say*		*Total*	
Variable	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*
	36	86	20	92	35	92		
	32	93	18	96	46	101		
	40	102	32	105	49	108		
	38	105	34	112	47	108		
	31	107	37	113	55	117		
	46	107	35	117	62	123		
	45	110	30	120	50	135		
			30	121	51	137		
					53	138		
					57	142		
Sample size, N_k	7		8		10		25	
Total	268	710	236	876	505	1201	1009	2787
Sum of squares	10,466	72,472	7,298	96,748	25,979	147,013	43,743	316,233
Sum of cross products	27,331		26,242		61,440		115,013	
Sample mean	38.2857	101.4286	29.5000	109.5000	50.5000	120 1000	40 3600	111.4800
Sample standard deviation	5.8513	8.7342	6 9282	10 8628	7 2763	17 5528	11 2171	15 1908
Regression slope	.3237		.4843		2847		4568	
Regression intercept	5.4576		−23.5266		16.3051		−10.5603	
Correlation coefficient	.4831		.7593		6869		6186	

IQ scores of the students and the resulting sample statistics, are shown in Table 20-5. The test scores represent the dependent variable of the study, the teaching methods represent the independent variable of the study, and the IQ measures represent what is termed the *covariate* of the study. Examination of the covariate scores shows that the classes were not comparable at the time training began since average IQs across the three classes are 101.4, 109.5, and 120 1. Because of these initial inequalities in intelligence, adjustments must be made in the dependent variable to correct for the inequalities that exist in the covariate before the hypothesis of no treatment effect can be tested by an analysis of covariance

This adjustment is achieved by determining, via the individual group regression equations, the values of the dependent variable, under the condition that all observations are measured at the grand average of the independent variable Once these adjusted dependent measures are made, an analysis of variance is performed upon them. Examination of Table 20-5 shows that the first subject in sample 1 has an IQ of 86 and a test score of 36 on the criterion measure. In addition, the grand average IQ of the entire set of students is given by $\bar{X} = 111.48$. As will be noted, the adjusting regression equation for the students in sample 1 is given by $\hat{Y}_X = 5.4576 + .3297X$. For an IQ of 86, the predicted test score is given by $\hat{Y}_X = 5.4576 + .3297(86) = 33.8$. Since this predicted score of 33.8 is below the actual score of 36, it follows that the earned score is $36 - 33.8 = 2.2$ points above average. If the IQ of this student had really been 111.48, then the predicted test score would be given by $\hat{Y}_X = 5.4576 + .3297(111.48) = 42.2$. So as to maintain the same relationship among the predicted score at $X = 86$, the actual earned score, the predicted score at $X = 111.48$, and its adjusted score, a 2.2 correction is added to this predicted score. Thus, the adjusted score is given by $\hat{Y}_X^A = 42.2 + 2.2 = 44.4$. Thus, if the student really had an IQ of 111.48, then his estimated earned score on the test would have equaled 44.4.

In terms of the geometry of Figure 20-8, it is seen that covariance adjustment is equivalent to projecting the earned score in a direction parallel to the regression line to the IQ score defined at $\bar{X}$. This parallel projection is performed for all pairs of observations and an analysis of variance is then performed on the adjusted scores.

In addition to the usual analysis-of-variance assumptions of normality, common variance, independence between samples, and independence within samples, one must assume in an analysis-of-covariance design that the regression lines for the dependent variable as produced from the covariate are parallel. Inspection of Figure 20-9, which is a graphic presentation of the regression lines over the observed range of IQ, suggests that in the hypothetical universes, the regressions are indeed parallel or close to being parallel. Fortunately, if the model involves only one covariate, a test of this hypothesis can be made with little difficulty by computing the statistics defined in Theorems 20-10 through 20-14, which are stated without proof.

For the one-sample regression model, the statistical test of $H_0\colon \beta = \beta_0$ is given by

$$t = \frac{\hat{\beta} - \beta_0}{\mathrm{SE}_{\beta_0}} \qquad \text{with } \nu = N - 1$$

Since $t_\nu^2 = F_{1,\nu}$, the test of H_0: $\beta = \beta_0$ can be performed by using

$$F = \frac{(\hat{\beta} - \beta_0)^2}{\mathrm{SE}_{\hat{\beta}_0}^2} = \frac{(N-1)\,S_X^2(\hat{\beta} - \beta_0)^2}{S_{Y.X}^2}$$

with $\nu_1 = 1$ and $\nu_2 = N - 1$. If $\beta_0 = 0$,

$$F = \frac{(N-1)\,S_X^2\hat{\beta}^2}{S_{Y.X}^2} = \frac{\mathrm{MS}(\hat{\beta})}{\mathrm{MSR}}$$

Since $\mathrm{MS}(\hat{\beta}) = \mathrm{SS}(\hat{\beta})/1$, it follows that

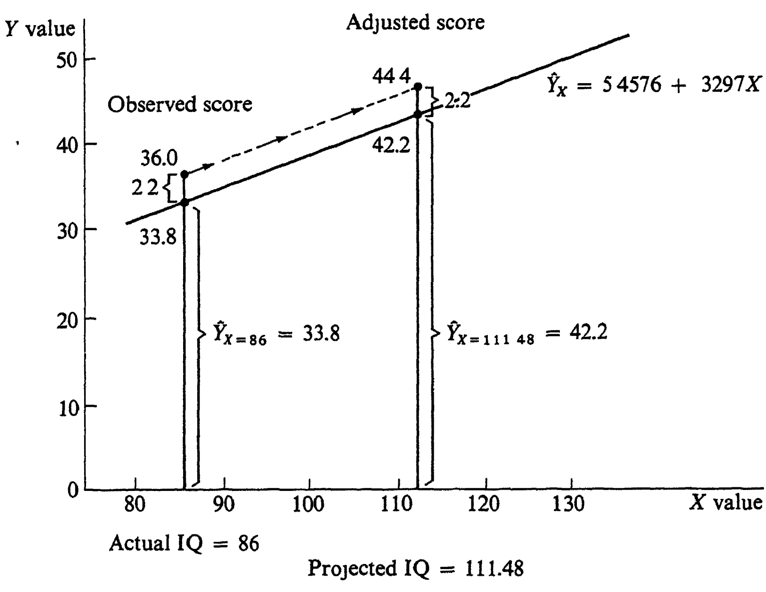
$$\mathrm{SS}(\hat{\beta}) = (N-1)\,S_X^2\hat{\beta}^2$$

Because the samples in the K-sample regression model are independent, the sum of squares across all samples is given by

$$\mathrm{SS}(\hat{\beta}_1) + \mathrm{SS}(\hat{\beta}_2) + \cdots + \mathrm{SS}(\hat{\beta}_K) = (N_1 - 1)\,S_{X1}^2\hat{\beta}_1^2 + (N_2 - 1)\,S_{X2}^2\hat{\beta}_2^2 + \cdots + (N_K - 1)\,S_{XK}^2\hat{\beta}_K^2$$

with $\nu = 1 + 1 + \cdots + 1 = K$. If $\beta_1 = \beta_2 = \cdots = \beta_K = \beta_0$, then one should adjust this sum of squares by removing the effects of the covariate upon the dependent measure. This adjustment is attained by estimating $\hat{\beta}_0$ and then subtracting from the unadjusted sum of squares the sum of squares associated with this parameter. When all β_k are equal, an unbiased estimate of β_0 is given by the formula of Theorem 20-10.

Figure 20-8. The geometry of covariance adjustment for the student with $Y = 36$ and $X = 86$.



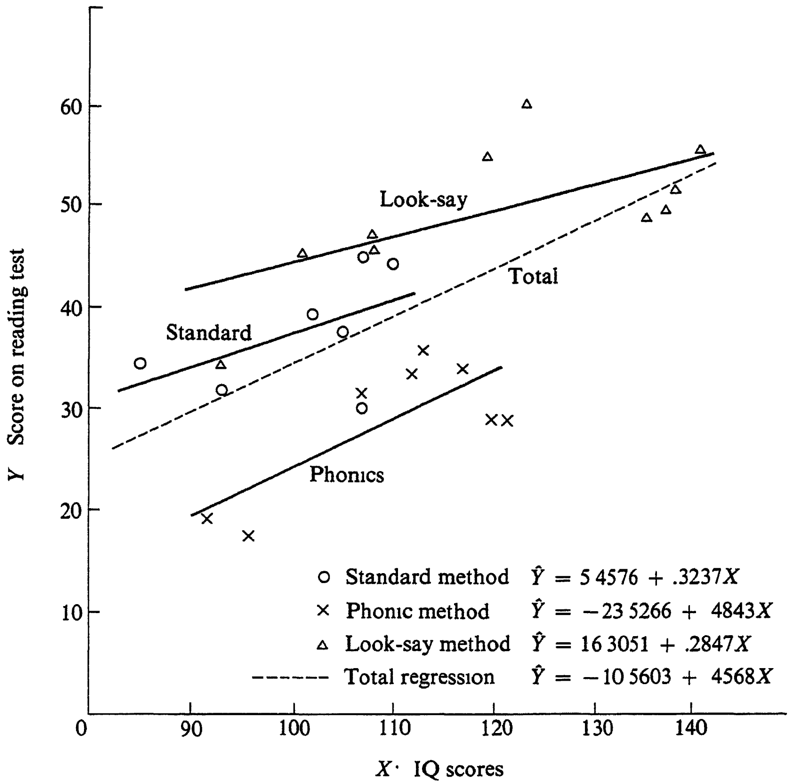


Figure 20-9. Scatter diagram for the analysis-of-covariance study of three reading instruction methods along with the estimated regression lines.

Theorem 20-10

An unbiased estimate of the slope of K regression lines that are known to be parallel is given by

$$\hat{\beta}_0 = \frac{(N_1 - 1) S_{X1}^2 \hat{\beta}_1 + (N_2 - 1) S_{X2}^2 \hat{\beta}_2 + \cdots + (N_K - 1) S_{XK}^2 \hat{\beta}_K}{(N_1 - 1) S_{X1}^2 + (N_2 - 1) S_{X2} + \cdots + (N_K - 1) S_{XK}^2}$$

Since this estimated value is based upon the within sample regression slopes, the K-sample analog to Theorem 20-9 is as follows.

Theorem 20-11

The sum of squares associated with the linear component of a K-sample regression model with parallel slopes is given by

$$\mathrm{SS}(\hat{\beta}_0) = (N - K)\,\mathrm{MSW}_X \hat{\beta}_0^2 = \hat{\beta}_0^2 \sum_{k=1}^{K} (N_k - 1) S_{Xk}^2$$

where MSW_X is the mean square within of the covariate. Thus, the sum of squares associated with $\beta_1 = \beta_2 = \cdots = \beta_K = \beta_0$ is given by the formula of Theorem 20-12.

Theorem 20-12

The sum of squares associated with the assumption of equal slopes in a K-sample regression model with parallel slopes is given by

$$\begin{aligned} SS(\hat{\beta}) &= SS(\hat{\beta}_1) + SS(\hat{\beta}_2) + \cdot + SS(\hat{\beta}_K) - SS(\hat{\beta}_0) \\ &= \sum_{k=1}^{K} (N_k - 1) S_{Xk}^2 \hat{\beta}_k^2 - \hat{\beta}_0^2 \sum_{k=1}^{K} (N_k - 1) S_{Xk}^2 \end{aligned}$$

with $\nu = K - 1$.

In the one-sample regression model, the sum of squares associated with the residual is given by

$$SSR = (N-1) S_Y^2 - (N-1) S_X^2 \hat{\beta}^2$$

Since the samples in the K-sample model are independent, the sum of squares associated with the residual across all samples is given by the following theorem.

Theorem 20-13

The sum of squares residual in a K-sample regression model is given by

$$\begin{aligned} SSR = [(N_1 - 1) S_{Y1}^2 - (N_1 - 1) S_{X1}^2 \hat{\beta}_1^2] + [(N_2 - 1) S_{Y2}^2 - (N_2 - 1) S_{X2}^2 \hat{\beta}_2^2] + \cdots \\ + [(N_K - 1) S_{YK}^2 - (N_K - 1) S_{XK}^2 \hat{\beta}_K^2] \end{aligned}$$

$$= \sum_{k=1}^{K} (N_k - 1) S_{YK}^2 - \sum_{k=1}^{K} (N_k - 1) S_{Xk}^2 \hat{\beta}_k^2$$

with $\nu = (N_1 - 2) + (N_2 - 2) + \cdots + (N_K - 2) = N - 2K$.

In the one-sample regression model, the sum of squares total is given by

$$\begin{aligned} SST &= SS(\hat{\beta}) + SSR \\ &= (N-1) S_X^2 \hat{\beta}^2 + [(N-1) S_Y^2 - (N-1) S_X^2 \hat{\beta}^2] \\ &= (N-1) S_Y^2 \end{aligned}$$

By direct analogy one has, in the K-sample regression mode, the following theorem.

Theorem 20-14

In the K-sample model, the sum of squares total is given by

$$\begin{aligned} \mathrm{SST} &= \mathrm{SS}(\hat{\beta}_0) + \mathrm{SS}(\hat{\beta}) + \mathrm{SSR} \\ &= (N-K)\,\mathrm{MSW}_X\,\hat{\beta}_0^2 + \left[\sum_{k=1}^{K}(N_k-1)\,S_{Yk}^2\hat{\beta}_k^2 - (N-K)\,\mathrm{MSW}_X\,\hat{\beta}_0^2\right] \\ &\qquad + \left[\sum_{k=1}^{K}(N_k-1)\,S_{Yk}^2 - \sum_{k=1}^{K}(N_k-1)\,S_{Xk}^2\,\hat{\beta}_k^2\right] \\ &= \sum_{k=1}^{K}(N_k-1)\,S_{Yk}^2 \\ &= (N-K)\,\mathrm{MSW}_Y \end{aligned}$$

with $\nu = 1 + (K-1) + (N-2K) = N-K$, where MSW_Y equals the mean square within for the dependent variable.

Table 20-6. Analysis-of-variance table for testing H_0 $\beta_1 = \beta_2 = \cdot = \beta_K = \beta_0$

Source	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F ratio*
Slope	1	$\mathrm{SS}(\hat{\beta}_0)$	$\mathrm{MS}(\hat{\beta}_0)$	$\dfrac{\mathrm{MS}(\hat{\beta}_0)}{\mathrm{MSR}}$
Parallelism	$K-1$	$\mathrm{SS}(\hat{\beta})$	$\mathrm{MS}(\hat{\beta})$	$\dfrac{\mathrm{MS}(\hat{\beta})}{\mathrm{MSR}}$
Residual	$N-2K$	SSR	MSR	
Total	$N-K$	SST		

With these sums of squares the analysis-of-variance table is as shown in Table 20-6. For the sample data,

$$\begin{aligned} \hat{\beta}_0 &= \frac{\sum_{k=1}^{3}(N_k-1)\,S_{Xk}^2\,\hat{\beta}_k}{\sum_{k=1}^{3}(N_k-1)\,S_{Xk}^2} \\ &= \frac{6(8.7342)^2(.3237) + 7(10.8628)^2(.4843) + 9(17.5528)^2(.2847)}{6(8.7342)^2 + 7(10.8628)^2 + 9(17.5528)^2} \\ &= \frac{1{,}337.6430}{4{,}056.6275} \\ &= .3297 \end{aligned}$$

$$
\begin{aligned}
\mathrm{SS}(\hat{\beta}_0) &= \hat{\beta}_0^2 \sum_{k=1}^{3} (N_k - 1)\, S_{Xk}^2 \\
&= (.3297)^2 (4{,}056.6275) \\
&= 440.9554 \\
\mathrm{SS}(\hat{\beta}) &= \sum_{k=1}^{3} (N_k - 1)\, S_{Xk}^2 \hat{\beta}_k^2 - \hat{\beta}_0^2 \sum_{k=1}^{3} (N_k - 1)\, S_{Xk}^2 \\
&= [6(8.7342)^2 (\ 3237)^2 + 7(10.8621)^2 (.4843)^2 \\
&\qquad + 9(17.5528)^2 (.2847)^2] - 440.9554 \\
&= 466.4427 - 440.9554 \\
&= 25.4873 \\
\mathrm{SSR} &= \sum_{k=1}^{K} (N_k - 1)\, S_{Yk}^2 - \sum_{k=1}^{K} (N_k - 1)\, S_{Xk}^2 \hat{\beta}_k^2 \\
&= [6(5.8513)^2 + 7(6.9282)^2 + 9(7.2763)^2] - 466.4427 \\
&= 1{,}071.9269 - 466.4427 \\
&= 551.4842 \\
\mathrm{SST} &= \mathrm{SS}(\hat{\beta}_0) + \mathrm{SS}(\hat{\beta}) + \mathrm{SSR} \\
&= 440.9554 + 25.4873 + 551.4842 \\
&= 1{,}071.9268
\end{aligned}
$$

Table 20-7. Analysis-of-variance table for testing H_0 $\beta_1 = \beta_2 = \beta_3 = \beta_0$ versus H_1 . H_0 is false.

Source	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F ratio*
Slope	1	440 9554	440 9554	15 19
Parallelism	2	25.4873	12 7437	44
Residual	19	551 4842	29 0255	
Total	22	1,017 9269		

The analysis-of-variance table for these data is shown in Table 20-7. Since $F = \mathrm{MS}(\hat{\beta})/\mathrm{MSR} = .44$ is less than $F_{2,19}(.95) = 3.52$, the hypothesis of equal slopes is not rejected. The test for the effects of the covariate upon the dependent variable is given by $F = \mathrm{MS}(\hat{\beta}_0)/\mathrm{MSR} = 15.19$. Since $F_{1,19}(.95) = 4.38$, the hypothesis of no effect is rejected. This means that an analysis of variance on the Y variable without an adjustment based on the X variable is not appropriate, since performance

on the criterion test is correlated with IQ. The average value of the correlation is given by

$$\bar{r} = \sqrt{\frac{\text{SS}(\hat{\beta}_0)}{\text{SST}}} = \sqrt{\frac{440.9554}{1{,}017.9269}} = \sqrt{.4332} = .66$$

When the hypothesis of equal slopes is rejected, one knows that the regression lines are not parallel and that the independent variable of the study interacts with the covariate so as to have differential effects upon the dependent or criterion variable. On the other hand, if the hypothesis of equal slopes is not rejected, then the regression lines are parallel, and it is known that the independent variable has a uniform effect upon the covariate over its entire range of values. However, it is still possible that the manipulated variable could affect one or more conditions of the study so as to displace the intercepts of one or more of the regression lines. Thus, once one has established that $H_0\colon \beta_1 = \beta_2 = \cdots = \beta_K = \beta_0$ is a valid statement, one will still want to know whether $H_0\colon \alpha_1 = \alpha_2 = \cdots = \alpha_K = \alpha_0$ is a valid statement. If this latter hypothesis is rejected, then it follows that the manipulated variable has differential effects upon the dependent variable. While one could test $H_0\colon \alpha_1 = \alpha_2 = \cdots = \alpha_K = \alpha_0$ versus $H_1 \cdot H_0$ is false directly, it is customary to test a different but related hypothesis.

Under the assumption of equal slopes, the deviation form for the kth regression equation is given by

$$\hat{Y}_{ik} = \bar{Y}_k + \hat{\beta}_0(X_{ik} - \bar{X}_k)$$

For an individual with covariate value of $X_{ik} = \bar{X}$, the predicted value of Y_{ik} is given by

$$\hat{Y}_{ik} = \bar{Y}_k + \hat{\beta}_0(\bar{X} - \bar{X}_k)$$

This predicted value is called the adjusted estimated value. If all individuals in the kth group had their covariate values equal to $\bar{X}$, then $\hat{Y}_{ik} = \hat{Y}_k^A$ would represent the adjusted mean value for the kth group. If it were true that all of the adjusted means were equal to the same common value, then it would be reasonable to conclude that there are no differences between the groups after the adjustment for different covariate values had been made. This hypothesis can be represented by

$$H_0\colon E(\hat{Y}_1^A) = E(\hat{Y}_2^A) = \cdots = E(\hat{Y}_K^A)$$

and can be tested by an analysis of variance of the adjusted means.

It is easy to see that a hypothesis of equal values for the intercepts is identical to a hypothesis of equal adjusted means by noting that

$$\begin{aligned} E(\hat{Y}_k^A) &= \mu_{Yk} + \beta_0(\mu_X - \mu_{Xk}) \\ &= (\mu_{Yk} - \hat{\beta}_0\,\mu_{Xk}) + \beta_0\,\mu_X \\ &= \alpha_k + \beta_0\,\mu_X \end{aligned}$$

Since $\beta_0 \mu_X$ is a constant term added to each $E(\hat{Y}_k^A)$, the only way the adjusted values could differ from one another is if the α_k, the regression intercepts, differ. Thus, a test of equal adjusted means is identical to a test of equal intercepts, provided that the regression lines are parallel.

In the one-way analysis of variance, SST, SSW, and SSB are computed directly from the raw data. In the one-way analysis of covariance this practice is not followed since the adjusted sum of squares between, SSB^A, is difficult to compute. To circumvent this difficulty, one computes SST^A and SSR^A and then determines SSB^A by subtraction. For these computations one uses Theorem 20-15, which is stated without proof.

Theorem 20-15

The sums of squares for the analysis of covariance are given by

1. $$\mathrm{SSR}^A = \sum_{k=1}^{K}(N_k - 1)\, S_{Yk}^2 - \hat{\beta}_0^2 \sum_{k=1}^{K}(N_k - 1)\, S_{Xk}^2$$
$$= \mathrm{SSW}_Y - \hat{\beta}_0^2\, \mathrm{SSW}_X$$

with $\nu_R^A = (N - K) - 1 = N - K - 1$, and

2. $$\mathrm{SST}^A = \sum_{k=1}^{K} \sum_{i=1}^{N_k}(y_{ik} - \bar{Y})^2 - \hat{\hat{\beta}}_0^2 \sum_{k=1}^{K} \sum_{i=1}^{N_k}(x_{ik} - \bar{X})^2$$
$$= (N - 1)\, S_Y^2 - \hat{\hat{\beta}}_0^2 (N - 1)\, S_X^2$$
$$= \mathrm{SST}_Y - \hat{\hat{\beta}}_0^2\, \mathrm{SST}_X$$

with $\nu_T^A = (N - 1) - 1 = N - 2$, and

3. $\mathrm{SSB}^A = \mathrm{SST}^A - \mathrm{SSR}^A$

with $\nu_B^A = (N - 2) - (N - K - 1) = (K - 1)$, and

$$\hat{\hat{\beta}}_0 = \frac{\sum_{k=1}^{K} \sum_{i=1}^{N_k}(x_{ik} - \bar{X})(y_{ik} - \bar{Y})}{\sum_{k=1}^{K} \sum_{i=1}^{N_k}(x_{ik} - \bar{X})^2}$$

$$= \frac{N\left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} x_{ik} y_{ik}\right) - \left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} x_{ik}\right)\left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} y_{ik}\right)}{\left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} x_{ik}^2\right) - \left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} x_{ik}\right)^2}$$

Table 20-8. The one-way analysis-of-covariance table.

Source	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F ratio*
Between adjusted means	$K-1$	SSB^A	MSB^A	$\dfrac{MSB^A}{MSR^A}$
Residual	$N-K-1$	SSR^A	MSR^A	
Total	$N-2$	SST^A		

With this formulation, the analysis-of-covariance table is given as shown in Table 20-8. For the sample data

$$\hat{\hat{\beta}}_0 = \frac{N\left(\sum_{k=1}^{K}\sum_{i=1}^{N_k} x_{ik}y_{ik}\right) - \left(\sum_{k=1}^{K}\sum_{i=1}^{N_k} x_{ik}\right)\left(\sum_{k=1}^{K}\sum_{i=1}^{N_k} y_{ik}\right)}{N\left(\sum_{k=1}^{K}\sum_{i=1}^{N_k} x_{ik}^2\right) - \left(\sum_{k=1}^{K}\sum_{i=1}^{N_k} x_{ik}\right)^2}$$

$$= \frac{25(115{,}013) - (2{,}787)(1{,}009)}{25(43{,}743) - (2{,}787)^2}$$

$$= .4568$$

$$SST^A = (N-1)\,S_Y^2 - \hat{\hat{\beta}}{}_0^2(N-1)\,S_X^2$$

$$= 24(11.2171)^2 - (.4568)^2(24)(15.1908)^2$$

$$= 1{,}864.0992$$

$$SSR^A = \sum_{k=1}^{K}(N_k-1)\,S_{Yk}^2 - \hat{\beta}_0^2 \sum_{k=1}^{K}(N_k-1)\,S_{Xk}^2$$

$$= 1{,}017.9269 - (.3297)^2(4{,}056.6275)$$

$$= 576.9175$$

$$SSB^A = SST^A - SSR^A$$

$$= 1{,}864.0992 - 576.9715$$

$$= 1{,}287.0377$$

With these data, the analysis-of-covariance table is as shown in Table 20-9. Since $F = 23.42$ is larger than $F_{2,21}(.95) = 3.47$, the hypothesis of equal adjusted expected values is rejected. To identify possible reasons for the rejection, one would now investigate contrasts in the adjusted means. The appropriate standard error is stated in Theorem 20-16.

Table 20-9. The analysis-of-covariance table for testing $H_0: E(\hat{Y}_1^A) = E(\hat{Y}_2^A) = E(\hat{Y}_3^A)$.

Source	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F ratio*
Between	2	1287 0377	643.5189	23.42
Within	21	576.9715	27.4748	
Total	23	1864.0092		

Theorem 20-16

For simple contrasts $\psi = E(\hat{Y}_{k_1}^A) - E(\hat{Y}_{k_2}^A)$, the squared standard error is given by

$$\mathrm{SE}_{\hat{\psi}}^2 = \mathrm{MSR}^A\left[\frac{1}{N_{k_1}} + \frac{1}{N_{k_2}} + \frac{(\bar{X}_{k_1} - \bar{X}_{k_2})^2}{(N-K)\,\mathrm{MSW}_X}\right]$$

While one could investigate complex contrasts, this practice is generally avoided because the standard errors are difficult to compute. For this example, $S = \sqrt{2(3.47)}$, $(N-1)\,\mathrm{MSW}_X = 4{,}056\ 6275$, and $\mathrm{MSR}^A = 27.4748$. The simple *post hoc* comparisons for this study are summarized in Table 20-10. For these comparisons, the adjusted means are given by

$$\hat{Y}_k^A = \bar{Y}_k + \hat{\beta}_0(\bar{X} - \bar{X}_k)$$

For the observed data,

$$\hat{Y}_1^A = 38.2857 + .3297(111.4800 - 101.4286) = 41.5997$$
$$\hat{Y}_2^A = 29.5000 + .3297(111.4800 - 109.5000) = 30.1528$$
$$\hat{Y}_3^A = 50.5000 + .3297(111.4800 - 120.1000) = 47.6580$$

According to the confidence-interval statements, the phonics method of training differs from the standard and look-say methods, while these latter two methods are not statistically different from one another.

Finally, it should be noted that when the hypothesis of equal slopes is rejected, then a different approach is taken to assess the effects of the manipulated independent variable upon the dependent variable. One possibility is to base the analysis on simple contrasts and partition the total probability of a type I error according to the procedure discussed in Section 14-9. If this is not satisfactory, then one can either seek the aid of a statistician or design expert or else one can refer to an advanced text that treats this problem.

20-12 TESTS AND CONFIDENCE INTERVALS FOR REGRESSION MODELS

The statistical tests and confidence-interval procedures for regression models are summarized in Table 20-11.

Table 20-10. Computation table for the confidence intervals of the simple contrasts for the three-condition, fourth-grade reading study.

Contrast ψ	*Estimate of contrast* $\hat{\psi}$	$(\bar{X}_{k_1} - \bar{X}_{k_2})$	$\frac{(\bar{X}_{k_1} - \bar{X}_{k_2})^2}{4056.6275}$	$SE^2_{\hat{\psi}}$	$\psi = \hat{\psi} \pm 2\ 63 SE_{\hat{\psi}}$	*Decision*
$\psi_1 = E(\hat{Y}^A_1) - E(\hat{Y}^A_2)$	11 4469	−8 0714	0161	7.8000	11 4469 ± 7.3456	Significant
$\psi_2 = E(\hat{Y}^A_1) - E(\hat{Y}^A_3)$	−6.0583	−18 6714	.0859	8.9331	−6.0583 ± 7 9110	Not significant
$\psi_3 = E(\hat{Y}^A_2) - E(\hat{Y}^A_3)$	−17.5052	−10.6000	.0277	6.9426	17.5052 ± 6.9169	Significant

Table 20-11. Confidence intervals and tests for regression models.

Case	*Hypotheses*	*Test statistic*	*Confidence interval*	*Assumptions*
25	$H_0: \beta = \beta_0$ $H_1: \beta \neq \beta_0$	$t = \dfrac{\hat{\beta} - \beta_0}{\mathrm{SE}_{\hat{\beta}}}$ $\mathrm{SE}^2_{\hat{\beta}} = \dfrac{S^2_{Y.X}}{\sum_{k=1}^{K}(x_k - \bar{X})^2}$ $\nu = N - 2$	$\beta = \hat{\beta} \pm t_\nu\left(\dfrac{\alpha}{2}\right)\mathrm{SE}_{\hat{\beta}}$	1. Independence between pairs 2. Regression is linear 3. Normality 4. $\sigma^2_{Y.X}$ is unknown
26	$H_0. \alpha = \alpha_0$ $H_1: \alpha \neq \alpha_0$	$t = \dfrac{\hat{\alpha} - \alpha_0}{\mathrm{SE}_{\hat{\alpha}}}$ $\mathrm{SE}^2_{\hat{\alpha}} = \sum_{k=1}^{K} \dfrac{x_k^2}{N}\mathrm{SE}^2_{\hat{\beta}}$ $\nu = N - 2$	$\alpha = \hat{\alpha} \pm t_\nu\left(\dfrac{\alpha}{2}\right)\mathrm{SE}_{\hat{\alpha}}$	1. Independence between pairs 2. Regression is linear 3 Normality 4. $\sigma^2_{Y.X}$ is unknown
27	$H_0 \cdot \sigma^2_{Y.X} = \sigma^2_0$ $H_1: \sigma^2_{Y\,X} \neq \sigma^2_0$	$\chi^2 = \dfrac{(N-2)\,S^2_{Y\,X}}{\sigma^2_{Y.X}}$ $\nu = N - 2$	$\dfrac{(N-2)\,S^2_{Y.X}}{\chi^2_\nu(1-\alpha/2)} < \sigma^2_{Y\,X} < \dfrac{(N-2)\,S^2_{Y.X}}{\chi^2_\nu(\alpha/2)}$	1. Independence between pairs 2. Regression is linear 3. Normality 4. $\sigma^2_{Y.X}$ is unknown
28	$H_0 \cdot \beta_1 = \beta_2 = \cdots = \beta_K$ $H_1: H_0$ is false	$F = \dfrac{\mathrm{MS}(\beta)}{\mathrm{MSR}}$ $\nu_1 = K - 1$ $\nu_2 = N - 2K$	$\beta_{k_1} - \beta_{k_2} = \hat{\beta}_{k_1} - \hat{\beta}_{k_2} \pm S\sqrt{\mathrm{SE}^2_{\hat{\beta}k_1} + \mathrm{SE}^2_{\hat{\beta}k_2}}$ where $S^2 = (K-1)\,F_{K-1,N-2K}(1-\alpha)$ $k_1, k_2 = 1, 2, \quad , K$	1 Independence between pairs 2. Regression is linear 3. Normality 4. $\sigma^2_{Y\,X}$ is unknown 5. Independence between samples
29	$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_K$ $H_1: H_0$ is false	$F = \dfrac{\mathrm{MSB}^A}{\mathrm{MSR}^A}$ $\nu_1 = K - 1$ $\nu_2 = N - K - 1$	$E(\bar{Y}^A_{k_1}) - E(\bar{Y}^A_{k_2}) = \bar{Y}^A_{k_1} - \bar{Y}^A_{k_2} \pm S\sqrt{\mathrm{SE}^2_{\hat{\psi}}}$ where $S^2 = (K-1)\,F_{K-1,N-K-1}(1-\alpha)$ $\mathrm{SE}^2_{\hat{\psi}} = \mathrm{MSR}^A\left[\dfrac{1}{N_{k_1}} + \dfrac{1}{N_{k_2}} + \dfrac{(\bar{X}_{k_1} - \bar{X}_{k_2})^2}{(N-1)\,\mathrm{MSW}_X}\right]$	1. Independence between pairs 2. Regression is linear 3 Normality 4. $\sigma^2_{Y.X}$ is unknown 5 Independence between samples

20-13 SUMMARY

While correlation and regression problems are frequently confused, the differences between them are easy to identify. In a regression problem, one is concerned with estimating the relationship that exists between an independent variable X and a dependent variable Y, where the independent variable is the one manipulated by the researcher or by nature prior to data collection and where the dependent variable is observed to assess the effect that the independent variable has upon the fluctuations in the dependent variable. In a correlation problem, there is no true independent or dependent variable; both variables are dependent in that they covary together. In a regression model, the set of X values corresponds to a set of parameters that define the conditions of the study. Their exact numerical values are specified before data are collected or else they are collected simultaneously.

Regression problems are usually K-sample problems while a correlation problem is a one-sample problem. In a regression problem, the parameters of interest are α, β, and $\sigma^2_{Y \cdot X}$, where α and β define the regression equation $Y_X = \alpha + \beta X$ and where $\sigma^2_{Y \cdot X}$ defines the variance about the regression line at each preselected value of X. In a correlation problem, the parameters of interest are μ_1, μ_2, σ_1^2, σ_2^2, and ρ, with the greatest interest centering on ρ, the correlation coefficient, or the parameter that measures the strength of the association between Y_1 and Y_2.

A regression problem involves one regression line while a correlation problem involves two. Since, in a correlation problem, it is assumed that the joint probability distribution of the two variables Y_1, Y_2 is bivariate normal, the regression lines must both be linear. In a regression problem, the regression line need not be linear; frequently it is monotonic If Y always increases as X increases, the relationship between X and Y is monotonic increasing, while if Y always decreases as X increases, the relationship between X and Y is monotonic decreasing.

The main purpose of a correlation study is to assess the strength of relationship between two variables as estimated by their sample correlation coefficient. On the other hand, a regression study focuses on the prediction of the Y variable from knowledge of the X variable.

Frequently, correlation studies are analyzed as though they were regression studies When this is done, it must be assumed that one of the variables is measured with no error. While this may not be correct, it is often assumed in order to simplify analysis. If the X variable represents a true random sample from the universe of X values, then over repeated experimentations or trials one can use the conditional X outcomes to generalize to the larger universe of values. However, the validity of this extension requires random sampling on the X variable.

In a regression model, it is customary to estimate α and β by their least square estimates. In deriving these estimates one chooses α and β so that they minimize the sum of the squared deviations of the y_i from their predicted values Y_{X_i}. Following this procedure, one finds that

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

and

$$\hat{\beta} = \frac{N\left(\sum_{i=1}^{N} x_i y_i\right) - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)}{N\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i\right)^2}$$

so that

$$\hat{Y}_X = \hat{\alpha} + \hat{\beta}X = \bar{Y} + \hat{\beta}(X - \bar{X})$$

Since these estimates are unbiased, it follows that the estimated predicted values of Y_X are also unbiased.

If it is assumed that the variances about the unknown regression line are equal, then one can estimate this common value once the α and β are estimated. This estimate is given by

$$S^2_{Y.X} = \frac{1}{N-2}\sum_{i=1}^{N}(y_i - \hat{Y}_{X_i})^2$$

which is algebraically equivalent to

$$S^2_{Y.X} = \frac{N-1}{N-2}[S^2_Y - \hat{\beta}^2 S^2_X]$$

This estimate is then used to estimate the sampling variances of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{Y}_X$. These estimates are given by

$$\text{SE}^2_{\hat{\alpha}} = \frac{\sum_{i=1}^{N} x_i^2}{N}\text{SE}^2_{\hat{\beta}}$$

$$\text{SE}^2_{\hat{\beta}} = \frac{N}{\Delta_X}S^2_{Y.X}$$

$$\text{SE}^2_{\hat{Y}_X} = \frac{S^2_{Y.X}}{N}\left[1 + \frac{N^2(X - \bar{X})^2}{\Delta_X}\right]$$

where

$$\Delta_X = N\left(\sum_{i=1}^{N} x_i^2\right) - \left(\sum_{i=1}^{N} x_i\right)^2$$

Thus, to test the hypotheses $H_{01}: \beta = \beta_0$, $H_{02}: \alpha = \alpha_0$, and $H_{03}: Y_{X_0} = \alpha + \beta X_0$, one uses

$$t = \frac{\hat{\beta} - \beta_0}{\text{SE}_{\hat{\beta}}}, \qquad t = \frac{\hat{\alpha} - \alpha_0}{\text{SE}_{\hat{\alpha}}}, \qquad \text{and} \qquad t = \frac{\hat{Y}_{X_0} - Y_{X_0}}{\text{SE}_{\hat{Y}_{X_0}}}$$

and rejects the tested hypothesis if $t < t_{N-2}(\alpha/2)$ or if $t > t_{N-2}(1 - \alpha/2)$.

In a similar fashion, for the $(1-\alpha)$ percent confidence intervals the unknown parameters are given by

$$\hat{\beta} - t_{N-2}\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\beta}} < \beta < \hat{\beta} + t_{N-2}\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\beta}}$$

$$\hat{\alpha} - t_{N-2}\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\alpha}} < \alpha < \hat{\alpha} + t_{N-2}\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{\alpha}}$$

$$\hat{Y}_{X_0} - t_{N-2}\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{Y}_{X_0}} < Y_{X_0} < \hat{Y}_{X_0} + t_{N-2}\left(\frac{\alpha}{2}\right)\mathrm{SE}_{\hat{Y}_{X_0}}$$

The tests of hypotheses and confidence intervals for $\sigma^2_{Y\,X}$ are based upon the χ^2 distribution with $\nu = N-2$. For this case, one has

$$\chi^2 = (N-2)\frac{S^2_{Y.X}}{\sigma^2_{Y.X}}$$

as a test statistic, and

$$\frac{(N-2)\,S^2_{Y.X}}{\chi^2_{N-2}(1-\alpha/2)} < \sigma^2_{Y\,X} < \frac{(N-2)\,S^2_{Y.X}}{\chi^2_{N-2}(\alpha/2)}$$

as the confidence-interval statement.

The assumptions for these tests and confidence interval procedures are:

1. Random samples from each of the K populations.
2. Common variance for each population criterion variable.
3. Normality for the dependent variable or large enough sample sizes so that the conditional sample means are approximately normal.
4. The values of the independent variable are assigned prior to data collection.
5. The conditional means are defined by a linear equation.

The last assumption can be modified. For example, if it is assumed that the conditional means are defined by a quadratic equation

$$Y_X = \alpha + \beta X + \gamma X^2$$

one can use least-square theory to estimate α, β, and γ. In this case the normal equations are given by

$$\sum_{i=1}^{N} y_i = \alpha N + \beta \sum_{i=1}^{N} x_i + \gamma \sum_{i=1}^{N} x_i^2$$

$$\sum_{i=1}^{N} x_i y_i = \alpha \sum_{i=1}^{N} x_i + \beta \sum_{i=1}^{N} x_i^2 + \gamma \sum_{i=1}^{N} x_i^3$$

$$\sum_{i=1}^{N} x_i^2 y_i = \alpha \sum_{i=1}^{N} x_i^2 + \beta \sum_{i=1}^{N} x_i^3 + \gamma \sum_{i=1}^{N} x_i^4$$

These equations can be solved by determinants to give $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$. Using the theorems of Chapters 6 and 7, we can determine the $\mathrm{SE}^2_{\hat{\alpha}}$, $\mathrm{SE}^2_{\hat{\beta}}$, and $\mathrm{SE}^2_{\hat{\gamma}}$, which can be used to generate confidence intervals and tests of hypotheses analogous to those of this chapter.

Sometimes the values of the X variable cannot be specified in advance. When this occurs one can still use the model and formulas presented in this chapter to estimate the regression line and to test hypotheses.

Finally, one can always study a regression model as an analysis-of-variance model since it involves only one random variable. For this formulation one has

$$\mathrm{SST} = \mathrm{SS}(\hat{\beta}) + \mathrm{SS}(D\hat{\beta}) + \mathrm{SSW}$$

with

$$(N-1) = 1 + (K-2) + (N-K)$$

where

$$\mathrm{SSB} = \sum_{k=1}^{K} N_k(\bar{Y}_k - \bar{Y})^2$$

$$\mathrm{SS}(\hat{\beta}) = (N-1)\,\hat{\beta}^2 S_X^2$$

$$\mathrm{SS}(D\hat{\beta}) = \mathrm{SSB} - \mathrm{SS}(\hat{\beta})$$

$$\mathrm{SSW} = \sum_{k=1}^{K} (N_k - 1)\, S_k^2$$

$$\mathrm{SST} = \sum_{k=1}^{K} \sum_{i=1}^{N_k} (y_{ik} - \bar{Y})^2$$

which can be summarized as shown in Table 20-12.

Finally, regression models can be used in comparative studies where it is known that the dependent variable can be influenced by variables other than the researcher's manipulated variable. In this case, one refers to the K-sample model as an analysis-of-covariance model. Under the assumption of parallel regression lines, the test of equal-valued intercepts corresponds to the analysis-of-variance hypothesis of equal population centers, provided that an adjustment is made for the effects that the covariate has upon the dependent measure. The test for equal slopes or parallelism of the regression lines is given by

$$F = \frac{\mathrm{MS}(\hat{\beta})}{\mathrm{MSR}}$$

where

1. $\mathrm{MS}(\hat{\beta}) = \dfrac{1}{K-1} \displaystyle\sum_{k=1}^{K} (N_k - 1)\, S_{Xk}^2(\hat{\beta}_k^2 - \hat{\beta}_0^2)$

Table 20-12. Regression model in the analysis-of-variance table.

Source of variance	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F ratio*
Between groups	$K-1$	SSB		
Linear	1	$\mathrm{SS}(\hat{\beta})$	$\mathrm{MS}(\hat{\beta})$	$\dfrac{\mathrm{MS}(\hat{\beta})}{\mathrm{MSW}}$
Deviation from linearity	$K-2$	$\mathrm{SS}(D\hat{\beta})$	$\mathrm{MS}(D\hat{\beta})$	$\dfrac{\mathrm{MS}(D\hat{\beta})}{\mathrm{MSW}}$
Within groups	$N-K$	SSW	MSW	
Total	$N-1$	SST		

2. $$\hat{\beta}_0 = \frac{\sum_{k=1}^{K} (N_k - 1)\, S_{Xk}^2 \hat{\beta}_k}{\sum_{k=1}^{K} (N_k - 1)\, S_{Xk}^2}$$

3. $$\mathrm{MSR} = \frac{1}{N-2K} \sum_{k=1}^{K} (N_k - 1)(S_{Yk}^2 - \hat{\beta}_k^2 S_{Xk}^2)$$

The test for equal adjusted means is given by

$$F = \frac{\mathrm{MSB}^A}{\mathrm{MSR}^A}$$

where

1. $$\mathrm{MSB}^A = \frac{1}{K-1}\,\mathrm{SSB}^A$$

2. $$\mathrm{SSB}^A = \mathrm{SST}^A - \mathrm{SSR}^A$$

3. $$\mathrm{SST}^A = \mathrm{SST}_Y - \hat{\hat{\beta}}_0^2\,\mathrm{SST}_X$$

4. $$\hat{\hat{\beta}}_0 = \frac{N\left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} x_{ik} y_{ik}\right) - \left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} x_{ik}\right)\left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} y_{ik}\right)}{N\left(\sum_{k=1}^{K} \sum_{i=1}^{K_k} x_{ik}^2\right) - \left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} x_{ik}\right)^2}$$

5. $$\mathrm{SSR}^A = \mathrm{SSW}_Y - \hat{\beta}_0^2\,\mathrm{SSW}_X$$

6. $$\mathrm{MSR}^A = \frac{1}{N-K-1}\,\mathrm{SSR}^A$$

EXERCISES

20-1. Give some examples from behavioral studies that show a monotonic increasing or decreasing relationship between two quantitative variables.

20-2. For the data of Table 18-3 is there any evidence that $\alpha_{B.G} = 0$ or that $\beta_{B.G} = 0$? What assumptions have you made?

20-3. Show that

(a) $S^2_{Y.X} = \dfrac{N-1}{N-2}[S^2_Y - \hat{\beta}^2 S^2_X]$

(b) $E(S^2_{Y.X}) = \sigma^2_{Y.X}$

(c) $\text{SS}(\hat{\beta}) = (N-1)\hat{\beta}^2 S^2_X$

***20-4.** Analyze the data of Exercise 11-2 as a regression model. Summarize the results in an analysis-of-variance table. Which is a better testing model for these data: $H_0: \mu_0 = \mu_1 = \mu_5$ or $H_0: Y_X = \alpha + \beta X$, where $X = 0, 1, 5$? Why?

***20-5.** If the pretest scores of Exercise 11-8 are treated as fixed parameters associated with individual students, estimate $Y_X = \alpha + \beta X$ and test the hypothesis $H_0: \beta = 0$ Exactly what is one testing under this model?

***20-6.** Do the data of Table 13-4 represent a correlation or regression study? Defend your choice.

20-7. In a study in which the effects of a new remedial reading program were being evaluated, the following statistics were generated:

Classroom 1 Old program		*Classroom 2 New program*		*Classroom 3 Old program*		*Classroom 4 New program*	
Y	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*
5 1	4.3	6.3	4.7	6.6	6.2	7.3	6.0
5.6	5 1	6.0	5.6	5.9	6.4	8 2	6 3
5 9	5.6	7.1	5.6	7.1	6.5	7 1	6.6
6.5	5 7	7 3	5.9	7 3	6 6	6.9	6 8
6 3	5 9	7.1	5.9	6.9	6.6	7 9	6.8
6 8	5 9	6 9	5.9	7 1	6.9	8.2	6.9
7 0	5 9			7 4	6.9		

The reported statistics represent the grade-equivalent reading level. Thus, student 1 began the study reading at a fourth-grade-and-three-months reading level. At the end of the study, this student was reading at a fifth-grade-and-one-month reading level Is there any reason to believe that the new program is any better than the old program? To answer this question, perform an analysis of covariance on the data

***20-8.** The data of Table 14-1 have been considered in this book as an analysis-of-variance model. Note that one can define the independent variable in terms of the penalty scores of $X = 0, \frac{1}{2}, 1, 2$. In this sense, it could be treated as a regression model Do this, and summarize your results in an analysis-of-variance table.

***20-9.** For the repeated-measures data of Table 19-3, one might expect a monotonic increasing relationship to exist between performance and trial number. Just as in an independent group design one can test for linear or curvilinear components to the relationship between performance and trial. For this analysis one would have

Source of variance	*d/f*	
Between trials	4	
Linear		1
Curvilinear		3
Between subjects	7	
Within	28	
Total	39	

Complete the analysis of variance for this design. Which is the better analysis, one that tests H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ or one that tests H_0: $Y_X = \alpha + \beta X$? Defend your answer.

***20-10.** In an analysis-of-variance design in which the differences in the independent variable are equally spaced, use can be made of Scheffé contrasts to study the form of the regression, provided that $N_1 = N_2 = \cdots = N_K$. When $K = 5$, the contrast that measures the linear component is given by

$$\hat{\psi}_{\text{linear}} = -2\bar{X}_1 - 1\bar{X}_2 + 0\bar{X}_3 + 1\bar{X}_4 + 2\bar{X}_5$$

If the confidence interval for this contrast does not include zero, it is concluded that the regression has a linear component. The contrast for a quadratic component is given by

$$\hat{\psi}_{\text{quad}} = 2\bar{X}_1 - 1\bar{X}_2 - 2\bar{X}_3 - 1\bar{X}_4 + 2\bar{X}_5$$

If the confidence interval for this contrast does not include zero, it is concluded that the regression has a quadratic component. Using the data of Exercise 20-9, test for a linear or a quadratic component. What is your conclusion? For related studies, one can use the coefficients of orthogonal polynomials shown in Table A-13, where the coefficients for these contrasts were selected.

APPENDIXES

LIST OF STATISTICAL TABLES

Title

Binomial Coefficients $\binom{N}{x}$

Critical Values of X for the Sign Test

Probabilities for Hypergeometric Probabilities and Fourfold Tables for $N \leqslant 15$ or $N_1 + N_2 \leqslant 15$

Cumulative Probabilities for the Normal Distribution with $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$

Table of Random Numbers

Table of Critical Values for Testing Significance of Outliers for T (One-Sided Test) when Standard Deviation is Calculated from the Complete Sample

Percentiles of the Chi-Square Distribution

Percentiles of Student's t Distribution

Percentiles of the F Distribution

Critical Values for Cochran's Test of Equal Variance

Percentage Points of the Studentized Range and Coefficients for Tukey's Method

Fisher's Z Transformation

Coefficients for Orthogonal Polynomials for Trend Analysis

Table A-1. Binomial coefficients $\binom{N}{x}$.

N	$\binom{N}{0}$	$\binom{N}{1}$	$\binom{N}{2}$	$\binom{N}{3}$	$\binom{N}{4}$	$\binom{N}{5}$	$\binom{N}{6}$	$\binom{N}{7}$	$\binom{N}{8}$	$\binom{N}{9}$	$\binom{N}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	792	924	792	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756

Table A-2. Critical values of X for the sign test

(Two-tail percentage points are given for the binomial for $p = .5$)

N	1%	5%	10%	25%	N	1%	5%	10%	25%
1					51	15	18	19	20
2					52	16	18	19	21
3				0	53	16	18	20	21
4				0	54	17	19	20	22
5			0	0	55	17	19	20	22
6		0	0	1	56	17	20	21	23
7		0	0	1	57	18	20	21	23
8	0	0	1	1	58	18	21	22	24
9	0	1	1	2	59	19	21	22	24
10	0	1	1	2	60	19	21	23	25
11	0	1	2	3	61	20	22	23	25
12	1	2	2	3	62	20	22	24	25
13	1	2	3	3	63	20	23	24	26
14	1	2	3	4	64	21	23	24	26
15	2	3	3	4	65	21	24	25	27
16	2	3	4	5	66	22	24	25	27
17	2	4	4	5	67	22	25	26	28
18	3	4	5	6	68	22	25	26	28
19	3	4	5	6	69	23	25	27	29
20	3	5	5	6	70	23	26	27	29
21	4	5	6	7	71	24	26	28	30
22	4	5	6	7	72	24	27	28	30
23	4	6	7	8	73	25	27	28	31
24	5	6	7	8	74	25	28	29	31
25	5	7	7	9	75	25	28	29	32
26	6	7	8	9	76	26	28	30	32
27	6	7	8	10	77	26	29	30	32
28	6	8	9	10	78	27	29	31	33
29	7	8	9	10	79	27	30	31	33
30	7	9	10	11	80	28	30	32	34
31	7	9	10	11	81	28	31	32	34
32	8	9	10	12	82	28	31	33	35
33	8	10	11	12	83	29	32	33	35
34	9	10	11	13	84	29	32	33	36
35	9	11	12	13	85	30	32	34	36
36	9	11	12	14	86	30	33	34	37
37	10	12	13	14	87	31	33	35	37
38	10	12	13	14	88	31	34	35	38
39	11	12	13	15	89	31	34	36	38
40	11	13	14	15	90	32	35	36	39
41	11	13	14	16	91	32	35	37	39
42	12	14	15	16	92	33	36	37	39
43	12	14	15	17	93	33	36	38	40
44	13	15	16	17	94	34	37	38	40
45	13	15	16	18	95	34	37	38	41
46	13	15	16	18	96	34	37	39	41
47	14	16	17	19	97	35	38	39	42
48	14	16	17	19	98	35	38	40	42
49	15	17	18	19	99	36	39	40	43
50	15	17	18	20	100	36	39	41	43

For values of N larger than 100, approximate values of r may be found by taking the nearest integer less than $(N - 1)/2 - k\sqrt{N+1}$, where k is 1.2879, 0.9800, 0.8224, 0.5752 for the 1, 5, 10, 25% values, respectively.

Table A-3. Probabilities for hypergeometric probabilities and fourfold tables for $N \leqslant 15$ or $N_1 + N_2 \leqslant 15$.

N	S_1	S_2	X	PROBABILITY Obs.	Other	Total	N	S_1	S_2	X	PROBABILITY Obs	Other	Total	N	S_1	S_2	X	PROBABILITY Obs.	Other	Total
2	1	1	0	500	500	1 000	7	3	3	0	114	029	143	9	2	3	1	583	417	1 000
2	1	1	1	500	500	1 000	7	3	3	1	628	372	1 000	9	2	3	2	083	000	083
3	1	1	0	667	333	1 000	7	3	3	2	372	114	486	9	2	4	0	278	167	444
3	1	1	1	333	000	333	7	3	3	3	029	000	029	9	2	4	1	722	278	1 000
4	1	1	0	750	250	1 000	8	1	1	0	875	125	1 000	9	2	4	2	167	000	167
4	1	1	1	250	000	250	8	1	1	1	125	000	125	9	3	3	0	238	226	464
4	1	2	0	500	500	1 000	8	1	2	0	750	250	1 000	9	3	3	1	1 000	000	1 000
4	1	2	1	500	500	1 000	8	1	2	1	250	000	250	9	3	3	2	226	238	464
4	2	2	0	167	167	333	8	1	3	0	625	375	1 000	9	3	3	3	012	000	012
4	2	2	1	1 000	000	1 000	8	1	3	1	375	000	375	9	3	4	0	012	048	060
4	2	2	2	167	167	333	8	1	4	0	500	500	1 000	9	3	4	1	488	512	1 000
5	1	1	0	800	200	1 000	8	1	4	1	500	500	1 000	9	3	4	2	405	012	417
5	1	1	1	200	000	200	8	2	2	0	536	464	1 000	9	3	4	3	048	000	048
5	1	2	0	.600	400	1 000	8	2	2	1	464	.536	1 000	9	4	4	0	040	008	048
5	1	2	1	400	000	400	8	2	2	2	035	000	035	9	4	4	1	357	167	524
5	2	2	0	300	100	400	8	2	3	0	357	107	464	9	4	4	2	643	357	1 000
5	2	2	1	.700	300	1 000	8	2	3	1	643	357	1 000	9	4	4	3	167	040	206
5	2	2	2	100	000	100	8	2	3	2	107	000	107	9	4	4	4	008	000	008
6	1	1	0	833	167	1 000	8	2	4	0	214	214	428	10	1	1	0	900	100	1 000
6	1	1	1	167	000	167	8	2	4	1	1 000	000	1 000	10	1	1	1	100	000	100
6	1	2	0	667	333	1 000	8	2	4	2	214	214	428	10	1	2	0	800	200	1 000
6	1	2	1	333	000	333	8	3	3	0	179	018	197	10	1	2	1	200	000	200
6	1	3	0	500	500	1 000	8	3	3	1	715	286	1 000	10	1	3	0	700	300	1 000
6	1	3	1	500	500	1 000	8	3	3	2	286	179	465	10	1	3	1	300	000	300
6	2	2	0	400	067	467	8	3	3	3	018	000	018	10	1	4	0	600	400	1 000
6	2	2	1	533	.467	1 000	8	3	4	0	071	071	143	10	1	4	1	.400	000	400
6	2	2	2	067	000	067	8	3	4	1	500	500	1 000	10	1	5	0	500	500	1 000
6	2	3	0	200	200	400	8	3	4	2	500	500	1 000	10	1	5	1	500	500	1 000
6	2	3	1	1 000	.000	1 000	8	3	4	3	071	071	143	10	2	2	0	622	378	1 000
6	2	3	2	200	200	400	8	4	4	0	014	014	.029	10	2	2	1	378	000	378
6	3	3	0	050	050	100	8	4	4	1	243	243	486	10	2	2	2	022	000	022
6	3	3	1	500	500	1 000	8	4	4	2	1 000	000	1 000	10	2	3	0	467	067	533
6	3	3	2	500	500	1 000	8	4	4	3	243	243	486	10	2	3	1	533	467	1.000
6	3	3	3	050	050	100	8	4	4	4	014	014	.029	10	2	3	2	067	000	067
7	1	1	0	857	143	1 000	9	1	1	0	889	111	1 000	10	2	4	0	333	133	467
7	1	1	1	143	000	143	9	1	1	1	111	000	.111	10	2	4	1	667	333	1 000
7	1	2	0	714	286	1 000	9	1	2	0	778	.222	1 000	10	2	4	2	133	.000	.133
7	1	2	1	.286	000	286	9	1	2	1	222	000	222	10	2	5	0	222	222	.444
7	1	3	0	571	429	1 000	9	1	3	0	667	.333	1 000	10	2	5	1	778	778	1 000
7	1	3	1	429	000	429	9	1	3	1	333	.000	333	10	2	5	2	222	222	444
7	2	2	0	476	524	1 000	9	1	4	0	556	.444	1 000	10	3	3	0	292	183	475
7	2	2	1	524	000	524	9	1	4	1	444	.000	444	10	3	3	1	.708	292	1 000
7	2	2	2	048	000	048	9	2	2	0	583	.417	1 000	10	3	3	2	183	000	183
7	2	3	0	286	143	429	9	2	2	1	417	.000	417	10	3	3	3	008	000	008
7	2	3	1	714	286	1 000	9	2	2	2	028	000	028	10	3	4	0	.167	033	200
7	2	3	2	143	000	143	9	2	3	0	.417	.083	500	10	3	4	1	667	333	1 000

Cumulative probabilities are given for deviations in the observed direction from equality and for deviation of the same size or greater in the opposite direction. The total probabilities can be used for two-tail tests. Tables are extracted from more extensive tables prepared by Donald Goyette and M. Ray Mickey, Health Sciences Computing Facility, UCLA In the probability columns "Obs." refers to the probability of a deviation as large or larger in the observed direction and "Other" refers to the probability of a deviation as large or larger in the opposite direction

X	—	S_1
—	—	—
S_2	—	N

S_1 is the smallest marginal total and S_2 is the next smallest; X is the frequency in the cell corresponding to the two smallest totals.

Table A-3. Probabilities for hypergeometric probabilities and fourfold tables for $N \leqslant 15$ or $N_1 + N_2 \leqslant 15$ (*Continued*).

				PROBABILITY							PROBABILITY							PROBABILITY		
N	S_1	S_2	X	*Obs.*	*Other*	*Total*	N	S_1	S_2	X	*Obs*	*Other*	*Total*	N	S_1	S_2	X	*Obs*	*Other*	*Total*
10	3	4	2	333	167	500	11	3	4	2	279	212	491	12	3	3	1	618	382	1 000
10	3	4	3	033	0	033	11	3	4	3	024	0	024	12	3	3	2	127	0	127
10	3	5	0	083	083	167	11	3	5	0	121	061	182	12	3	3	3	005	0	005
10	3	5	1	500	500	1 000	11	3	5	1	576	424	1 000	12	3	4	0	255	236	491
10	3	5	2	500	500	1 000	11	3	5	2	424	121	545	12	3	4	1	764	745	1 000
10	3	5	3	083	083	167	11	3	5	3	061	0	061	12	3	4	2	236	255	491
10	4	4	0	071	005	076	11	4	4	0	106	088	194	12	3	4	3	018	0	018
10	4	4	1	452	119	571	11	4	4	1	530	470	1 000	12	3	5	0	159	045	205
10	4	4	2	548	452	1 000	11	4	4	2	470	106	576	12	3	5	1	636	364	1 000
10	4	4	3	119	071	190	11	4	4	3	088	0	088	12	3	5	2	364	159	523
10	4	4	4	005	0	005	11	4	4	4	003	0	003	12	3	5	3	045	0	045
10	4	5	0	024	024	048	11	4	5	0	045	015	061	12	3	6	0	091	091	182
10	4	5	1	262	262	524	11	4	5	1	348	197	545	12	3	6	1	500	500	1 000
10	4	5	2	738	738	1 000	11	4	5	2	652	348	1 000	12	3	6	2	500	500	1 000
10	4	5	3	262	262	524	11	4	5	3	197	045	242	12	3	6	3	091	091	182
10	4	5	4	024	024	048	11	4	5	4	015	0	015	12	4	4	0	141	067	208
10	5	5	0	004	004	008	11	5	5	0	013	002	015	12	4	4	1	594	406	1 000
10	5	5	1	103	103	206	11	5	5	1	175	067	242	12	4	4	2	406	141	547
10	5	5	2	500	500	1 000	11	5	5	2	608	392	1 000	12	4	4	3	067	0	067
10	5	5	3	500	500	1 000	11	5	5	3	392	175	567	12	4	4	4	002	0	002
10	5	5	4	103	103	206	11	5	5	4	067	013	080	12	4	5	0	071	010	081
10	5	5	5	004	004	008	11	5	5	5	002	0	002	12	4	5	1	424	152	576
11	1	1	0	909	091	1 000	12	1	1	0	917	083	1 000	12	4	5	2	576	424	1 000
11	1	1	1	091	0	091	12	1	1	1	083	0	083	12	4	5	3	152	071	222
11	1	2	0	818	182	1 000	12	1	2	0	833	167	1 000	12	4	5	4	010	0	010
11	1	2	1	182	0	182	12	1	2	1	167	0	167	12	4	6	0	030	030	061
11	1	3	0	727	273	1 000	12	1	3	0	750	250	1 000	12	4	6	1	273	273	545
11	1	3	1	273	0	273	12	1	3	1	250	0	250	12	4	6	2	727	727	1 000
11	1	4	0	636	364	1 000	12	1	4	0	667	333	1 000	12	4	6	3	273	273	545
11	1	4	1	364	0	364	12	1	4	1	333	0	333	12	4	6	4	030	030	061
11	1	5	0	545	455	1 000	12	1	5	0	583	417	1 000	12	5	5	0	027	001	028
11	1	5	1	455	0	455	12	1	5	1	417	0	417	12	5	5	1	247	045	293
11	2	2	0	655	345	1 000	12	1	6	0	500	500	1 000	12	5	5	2	689	311	1 000
11	2	2	1	345	0	345	12	1	6	1	500	500	1 000	12	5	5	3	311	247	558
11	2	2	2	018	0	018	12	2	2	0	682	318	1 000	12	5	5	4	045	027	072
11	2	3	0	509	055	564	12	2	2	1	318	0	318	12	5	5	5	001	0	001
11	2	3	1	491	509	1 000	12	2	2	2	015	0	015	12	5	6	0	008	008	015
11	2	3	2	055	0	055	12	2	3	0	545	455	1 000	12	5	6	1	121	121	242
11	2	4	0	382	109	491	12	2	3	1	455	545	1 000	12	5	6	2	500	500	1 000
11	2	4	1	618	382	1 000	12	2	3	2	045	0	045	12	5	6	3	500	500	1 000
11	2	4	2	109	0	109	12	2	4	0	424	091	515	12	5	6	4	121	121	242
11	2	5	0	273	182	455	12	2	4	1	576	424	1 000	12	5	6	5	008	008	015
11	2	5	1	727	273	1 000	12	2	4	2	091	0	091	12	6	6	0	001	001	002
11	2	5	2	.182	0	182	12	2	5	0	318	152	470	12	6	6	1	040	040	080
11	3	3	0	339	152	491	12	2	5	1	682	318	1 000	12	6	6	2	284	284	567
11	3	3	1	661	339	1 000	12	2	5	2	152	0	152	12	6	6	3	1 000	000	1 000
11	3	3	2	152	0	152	12	2	6	0	227	227	455	12	6	6	4	284	284	567
11	3	3	3	006	0	006	12	2	6	1	773	773	1 000	12	6	6	5	040	040	080
11	3	4	0	212	024	236	12	2	6	2	227	227	455	12	6	6	6	001	001	.002
11	3	4	1	721	279	1 000	12	3	3	0	382	127	509	13	1	1	0	923	077	1 000

Table A-3. Probabilities for hypergeometric probabilities and fourfold tables for $N \leqslant 15$ or $N_1 + N_2 \leqslant 15$ (*Continued*).

				PROBABILITY							PROBABILITY							PROBABILITY		
N	S_1	S_2	X	*Obs*	*Other*	*Total*	N	S_1	S_2	X	*Obs.*	*Other*	*Total*	N	S_1	S_2	X	*Obs.*	*Other*	*Total*
13	1	1	1	077	0	077	13	4	5	3	119	098	217	14	2	5	1	604	396	1 000
13	1	2	0	846	154	1 000	13	4	5	4	007	0	007	14	2	5	2	110	0	110
13	1	2	1	154	0	154	13	4	6	0	049	021	070	14	2	6	0	308	165	473
13	1	3	0	769	231	1 000	13	4	6	1	343	217	559	14	2	6	1	692	308	1 000
13	1	3	1	231	0	231	13	4	6	2	657	343	1 000	14	2	6	2	165	0	165
13	1	4	0	692	308	1 000	13	4	6	3	217	049	266	14	2	7	0	.231	231	462
13	1	4	1	308	0	308	13	4	6	4	021	0	021	14	2	7	1	769	769	1 000
13	1	5	0	615	385	1 000	13	5	5	0	044	032	075	14	2	7	2	231	231	462
13	1	5	1	385	0	385	13	5	5	1	315	249	565	14	3	3	0	453	093	547
13	1	6	0	538	462	1 000	13	5	5	2	685	315	1 000	14	3	3	1	547	453	1 000
13	1	6	1	462	0	462	13	5	5	3	249	044	293	14	3	3	2	093	0	093
13	2	2	0	705	295	1 000	13	5	5	4	032	0	032	14	3	3	3	003	0	003
13	2	2	1	295	0	295	13	5	5	5	001	0	001	14	3	4	0	330	176	505
13	2	2	2	013	0	013	13	5	6	0	016	005	021	14	3	4	1	670	330	1 000
13	2	3	0	577	423	1 000	13	5	6	1	179	086	266	14	3	4	2	176	0	176
13	2	3	1	423	0	423	13	5	6	2	587	413	1 000	14	3	4	3	011	0	011
13	2	3	2	038	0	038	13	5	6	3	413	179	592	14	3	5	0	231	027	258
13	2	4	0	462	077	538	13	5	6	4	086	016	103	14	3	5	1	.725	275	1 000
13	2	4	1	538	462	1 000	13	5	6	5	005	0	005	14	3	5	2	275	231	505
13	2	4	2	077	0	077	13	6	6	0	004	001	005	14	3	5	3	.027	0	027
13	2	5	0	359	128	487	13	6	6	1	078	025	103	14	3	6	0	154	055	209
13	2	5	1	641	359	1 000	13	6	6	2	383	209	592	14	3	6	1	615	385	1 000
13	2	5	2	128	0	128	13	6	6	3	617	383	1 000	14	3	6	2	385	154	538
13	2	6	0	269	192	462	13	6	6	4	209	078	286	14	3	6	3	055	0	055
13	2	6	1	731	269	1 000	13	6	6	5	025	004	029	14	3	7	0	096	096	192
13	2	6	2	192	0	192	13	6	6	6	001	0	001	14	3	7	1	500	500	1 000
13	3	3	0	420	108	528	14	1	1	0	929	071	1 000	14	3	7	2	500	500	1 000
13	3	3	1	580	420	1 000	14	1	1	1	071	0	071	14	3	7	3	096	096	192
13	3	3	2	108	0	108	14	1	2	0	857	143	1 000	13	4	4	0	210	041	251
13	3	3	3	003	0	003	14	1	2	1	143	0	143	14	4	4	1	689	311	1 000
13	3	4	0	294	203	497	14	1	3	0	786	214	1 000	14	4	4	2	311	210	520
13	3	4	1	706	294	1 000	14	1	3	1	214	0	214	14	4	4	3	041	0	041
13	3	4	2	203	0	203	14	1	4	0	714	286	1 000	14	4	4	4	001	0	.001
13	3	4	3	014	0	014	14	1	4	1	286	0	286	14	4	5	0	126	095	221
13	3	5	0	196	035	231	14	1	5	0	643	357	1 000	14	4	5	1	545	455	1 000
13	3	5	1	685	315	1 000	14	1	5	1	357	0	357	14	4	5	2	455	126	580
13	3	5	2	315	196	510	14	1	6	0	571	429	1 000	14	4	5	3	095	0	.095
13	3	5	3	035	0	035	14	1	6	1	429	0	429	14	4	5	4	005	0	005
13	3	6	0	122	070	192	14	1	7	0	500	500	1 000	14	4	6	0	070	.015	085
13	3	6	1	563	437	1 000	14	1	7	1	500	500	1 000	14	4	6	1	406	175	580
13	3	6	2	437	.122	559	14	2	2	0	725	275	1 000	14	4	6	2	594	406	1 000
13	3	6	3	070	0	070	14	2	2	1	275	0	275	14	4	6	3	175	070	245
13	4	4	0	176	052	228	14	2	2	2	011	0	.011	14	4	6	4	015	0	015
13	4	4	1	646	354	1 000	14	2	3	0	604	396	1 000	14	4	7	0	035	035	070
13	4	4	2	354	176	530	14	2	3	1	396	0	396	14	4	7	1	280	280	559
13	4	4	3	052	0	052	14	2	3	2	033	0	033	14	4	7	2	720	720	1 000
13	4	4	4	001	0	001	14	2	4	0	495	.066	560	14	4	7	3	280	280	559
13	4	5	0	098	.007	105	14	2	4	1	505	495	1 000	14	4	7	4	035	035	.070
13	4	5	1	490	.119	608	14	2	4	2	066	0	066	14	5	5	0	063	023	086
13	4	5	2	.510	.490	1 000	14	2	5	0	.396	110	505	14	5	5	1	.378	.203	580

Table A-3. Probabilities for hypergeometric probabilities and fourfold tables for $N \leqslant 15$ or $N_1 + N_2 \leqslant 15$ (*Continued*)

				PROBABILITY							PROBABILITY							PROBABILITY		
N	S_1	S_2	X	*Obs.*	*Other*	*Total*	N	S_1	S_2	X	*Obs*	*Other*	*Total*	N	S_1	S_2	X	*Obs.*	*Other*	*Total*
14	5	5	2	622	378	1 000	15	1	7	0	533	467	1 000	15	4	6	0	092	011	103
14	5	5	3	203	063	266	15	1	7	1	467	0	467	15	4	6	1	462	143	604
14	5	5	4	023	0	023	15	2	2	0	743	257	1 000	15	4	6	2	538	462	1 000
14	5	5	5	000	0	000	15	2	2	1	257	0	257	15	4	6	3	143	092	235
14	5	6	0	028	003	031	15	2	2	2	010	0	010	15	4	6	4	011	0	011
14	5	6	1	238	063	301	15	2	3	0	629	371	1 000	15	4	7	0	051	026	077
14	5	6	2	657	343	1 000	15	2	3	1	371	0	371	15	4	7	1	338	231	569
14	5	6	3	343	238	580	15	2	3	2	029	0	029	15	4	7	2	662	338	1 000
14	5	6	4	063	028	091	15	2	4	0	524	057	581	15	4	7	3	231	051	282
14	5	6	5	003	0	003	15	2	4	1	476	524	1 000	15	4	7	4	026	0	026
14	5	7	0	010	010	021	15	2	4	2	057	0	057	15	5	5	0	084	017	101
14	5	7	1	133	133	266	15	2	5	0	429	095	524	15	5	5	1	434	167	600
14	5	7	2	500	500	1 000	15	2	5	1	571	429	1 000	15	5	5	2	566	434	1 000
14	5	7	3	500	500	1 000	15	2	5	2	095	0	095	15	5	5	3	167	084	251
14	5	7	4	133	133	266	15	2	6	0	343	143	486	15	5	5	4	017	0	017
14	5	7	5	010	010	021	15	2	6	1	657	343	1 000	15	5	5	5	000	0	000
14	6	6	0	009	000	010	15	2	6	2	143	0	143	15	5	6	0	042	047	089
14	6	6	1	121	016	138	15	2	7	0	267	200	467	15	5	6	1	294	287	580
14	6	6	2	471	156	627	15	2	7	1	733	267	1 000	15	5	6	2	713	706	1 000
14	6	6	3	529	471	1 000	15	2	7	2	200	0	200	15	5	6	3	287	294	580
14	6	6	4	156	121	277	15	3	3	0	484	081	565	15	5	6	4	047	042	089
14	6	6	5	016	009	026	15	3	3	1	516	484	1 000	15	5	6	5	002	0	002
14	6	6	6	000	0	000	15	3	3	2	081	0	081	15	5	7	0	019	007	026
14	6	7	0	002	002	005	15	3	3	3	002	0	002	15	5	7	1	182	100	282
14	6	7	1	051	051	103	15	3	4	0	363	154	516	15	5	7	2	573	427	1 000
14	6	7	2	296	296	592	15	3	4	1	637	363	1 000	15	5	7	3	427	182	608
14	6	7	3	704	704	1 000	15	3	4	2	154	0	154	15	5	7	4	100	019	119
14	6	7	4	296	296	592	15	3	4	3	009	0	009	15	5	7	5	007	0	007
14	6	7	5	051	051	103	15	3	5	0	264	242	505	15	6	6	0	017	011	028
14	6	7	6	002	002	005	15	3	5	1	758	736	1 000	15	6	6	1	168	119	287
14	7	7	0	000	000	001	15	3	5	2	242	264	505	15	6	6	2	545	455	1 000
14	7	7	1	015	015	029	15	3	5	3	022	0	022	15	6	6	3	455	168	622
14	7	7	2	143	143	286	15	3	6	0	185	044	229	15	6	6	4	119	017	136
14	7	7	3	500	500	1 000	15	3	6	1	659	341	1 000	15	6	6	5	011	0	011
14	7	7	4	.500	500	1 000	15	3	6	2	341	185	525	15	6	6	6	000	0	000
14	7	7	5	143	143	286	15	3	6	3	044	0	044	15	6	7	0	006	001	007
14	7	7	6	015	015	029	15	3	7	0	123	077	200	15	6	7	1	084	035	119
14	7	7	7	000	000	001	15	3	7	1	554	446	1 000	15	6	7	2	378	231	608
15	1	1	0	933	067	1 000	15	3	7	2	446	123	569	15	6	7	3	622	378	1 000
15	1	1	1	067	0	067	15	3	7	3	077	0	077	15	6	7	4	231	084	315
15	1	2	0	867	133	1 000	15	4	4	0	242	033	275	15	6	7	5	035	006	041
15	1	2	1	133	0	133	15	4	4	1	725	275	1 000	15	6	7	6	001	0	001
15	1	3	0	800	200	1 000	15	4	4	2	275	242	516	15	7	7	0	001	000	001
15	1	3	1	200	0	200	15	4	4	3	033	0	033	15	7	7	1	032	009	.041
15	1	4	0	733	267	1 000	15	4	4	4	001	0	001	15	7	7	2	214	100	315
15	1	4	1	267	0	267	15	4	5	0	154	077	231	15	7	7	3	595	405	1 000
15	1	5	0	667	333	1 000	15	4	5	1	593	407	1 000	15	7	7	4	405	214	619
15	1	5	1	.333	0	333	15	4	5	2	407	.154	560	15	7	7	5	100	032	.132
15	1	6	0	600	400	1 000	15	4	5	3	077	0	077	15	7	7	6	009	001	.010
15	1	6	1	400	0	400	15	4	5	4	004	0	004	15	7	7	7	000	0	000

Table A-4 Cumulative probabilities for the normal distribution with $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$.

z	X	Area	z	X	Area
1.05	$\mu + 1.05\sigma$	.8531	−4.265	$\mu - 4.265\sigma$	.00001
1.10	$\mu + 1.10\sigma$	.8643	−3.719	$\mu - 3.719\sigma$	.0001
1.15	$\mu + 1.15\sigma$	.8749	−3.090	$\mu - 3.090\sigma$	.001
1.20	$\mu + 1.20\sigma$	.8849	−2.576	$\mu - 2.576\sigma$	.005
1.25	$\mu + 1.25\sigma$	.8944	−2.326	$\mu - 2.326\sigma$	.01
1.30	$\mu + 1.30\sigma$	.9032	−2.054	$\mu - 2.054\sigma$	.02
1.35	$\mu + 1.35\sigma$	.9115	−1.960	$\mu - 1.960\sigma$	.025
1.40	$\mu + 1.40\sigma$	.9192	−1.881	$\mu - 1.881\sigma$	.03
1.45	$\mu + 1.45\sigma$	.9265	−1.751	$\mu - 1.751\sigma$	.04
1.50	$\mu + 1.50\sigma$	.9332	−1.645	$\mu - 1.645\sigma$	.05
1.55	$\mu + 1.55\sigma$	.9394	−1.555	$\mu - 1.555\sigma$	.06
1.60	$\mu + 1.60\sigma$	.9452	−1.476	$\mu - 1.476\sigma$	.07
1.65	$\mu + 1.65\sigma$	.9505	−1.405	$\mu - 1.405\sigma$	.08
1.70	$\mu + 1.70\sigma$	.9554	−1.341	$\mu - 1.341\sigma$	.09
1.75	$\mu + 1.75\sigma$	.9599	−1.282	$\mu - 1.282\sigma$	.10
1.80	$\mu + 1.80\sigma$	.9641	−1.036	$\mu - 1.036\sigma$	.15
1.85	$\mu + 1.85\sigma$	.9678	−.842	$\mu - .842\sigma$	.20
1.90	$\mu + 1.90\sigma$	.9713	−.674	$\mu - .674\sigma$	.25
1.95	$\mu + 1.95\sigma$	.9744	−.524	$\mu - .524\sigma$	.30
2.00	$\mu + 2.00\sigma$	.9772	−.385	$\mu - .385\sigma$	.35
2.05	$\mu + 2.05\sigma$	.9798	−.253	$\mu - .253\sigma$	.40
2.10	$\mu + 2.10\sigma$	.9821	−.126	$\mu - .126\sigma$	.45
2.15	$\mu + 2.15\sigma$	.9842	0	μ	.50
2.20	$\mu + 2.20\sigma$	.9861	.126	$\mu + .126\sigma$	.55
2.25	$\mu + 2.25\sigma$	.9878	.253	$\mu + .253\sigma$	.60
2.30	$\mu + 2.30\sigma$	.9893	.385	$\mu + .385\sigma$	.65
2.35	$\mu + 2.35\sigma$	.9906	.524	$\mu + .524\sigma$	.70
2.40	$\mu + 2.40\sigma$	.9918	.674	$\mu + .674\sigma$	.75
2.45	$\mu + 2.45\sigma$	.9929	.842	$\mu + .842\sigma$	.80
2.50	$\mu + 2.50\sigma$	.9938	1.036	$\mu + 1.036\sigma$	.85
2.55	$\mu + 2.55\sigma$	.9946	1.282	$\mu + 1.282\sigma$	.90
2.60	$\mu + 2.60\sigma$	.9953	1.341	$\mu + 1.341\sigma$	.91
2.65	$\mu + 2.65\sigma$	.9960	1.405	$\mu + 1.405\sigma$	.92
2.70	$\mu + 2.70\sigma$	.9965	1.476	$\mu + 1.476\sigma$	.93
2.75	$\mu + 2.75\sigma$	.9970	1.555	$\mu + 1.555\sigma$	.94
2.80	$\mu + 2.80\sigma$	.9974	1.645	$\mu + 1.645\sigma$	.95
2.85	$\mu + 2.85\sigma$	.9978	1.751	$\mu + 1.751\sigma$	.96
2.90	$\mu + 2.90\sigma$	.9981	1.881	$\mu + 1.881\sigma$	.97
2.95	$\mu + 2.95\sigma$	.9984	1.960	$\mu + 1.960\sigma$	.975
3.00	$\mu + 3.00\sigma$	.9987	2.054	$\mu + 2.054\sigma$	.98
3.05	$\mu + 3.05\sigma$	.9989	2.326	$\mu + 2.326\sigma$	.99
3.10	$\mu + 3.10\sigma$	.9990	2.576	$\mu + 2.576\sigma$	.995
3.15	$\mu + 3.15\sigma$	.9992	3.090	$\mu + 3.090\sigma$	.999
3.20	$\mu + 3.20\sigma$	.9993	3.719	$\mu + 3.719\sigma$	.9999
3.25	$\mu + 3.25\sigma$	.9994	4.265	$\mu + 4.265\sigma$	.99999

Table A-4. Cumulative probabilities for the normal distribution with $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$ (*Continued*)

z	X	Area	z	X	Area
−3 25	$\mu - 3\ 25\sigma$	.0006	−1 00	$\mu - 1.00\sigma$	.1587
−3 20	$\mu - 3\ 20\sigma$	.0007	− 95	$\mu - .95\sigma$	.1711
−3 15	$\mu - 3\ 15\sigma$	.0008	− 90	$\mu - .90\sigma$	.1841
−3.10	$\mu - 3.10\sigma$	.0010	− 85	$\mu - 85\sigma$	.1977
−3 05	$\mu - 3.05\sigma$	0011	− 80	$\mu - 80\sigma$	.2119
−3 00	$\mu - 3\ 00\sigma$	.0013	− 75	$\mu - .75\sigma$	.2266
−2 95	$\mu - 2\ 95\sigma$	0016	− 70	$\mu - 70\sigma$	.2420
−2 90	$\mu - 2\ 90\sigma$	0019	− 65	$\mu - .65\sigma$	.2578
−2 85	$\mu - 2\ 85\sigma$	0022	− 60	$\mu - 60\sigma$	.2743
−2 80	$\mu - 2\ 80\sigma$	0026	− 55	$\mu - 55\sigma$	.2912
−2.75	$\mu - 2\ 75\sigma$	0030	− 50	$\mu - .50\sigma$	.3085
−2 70	$\mu - 2.70\sigma$	0035	− 45	$\mu - .45\sigma$	.3264
−2 65	$\mu - 2\ 65\sigma$	0040	− 40	$\mu - .40\sigma$	.3446
−2 60	$\mu - 2\ 60\sigma$	0047	− .35	$\mu - .35\sigma$	.3632
−2 55	$\mu - 2.55\sigma$	0054	− 30	$\mu - 30\sigma$	.3821
−2 50	$\mu - 2\ 50\sigma$	0062	− 25	$\mu - .25\sigma$	.4013
−2 45	$\mu - 2\ 45\sigma$	0071	− .20	$\mu - .20\sigma$	.4207
−2.40	$\mu - 2\ 40\sigma$	0082	− .15	$\mu - 15\sigma$	.4404
−2 35	$\mu - 2\ 35\sigma$	0094	− .10	$\mu - 10\sigma$	.4602
−2 30	$\mu - 2\ 30\sigma$	0107	− .05	$\mu - .05\sigma$	.4801
−2 25	$\mu - 2.25\sigma$	0122			
−2 20	$\mu - 2.20\sigma$	0139			
−2 15	$\mu - 2\ 15\sigma$	0158	.00	μ	.5000
−2 10	$\mu - 2\ 10\sigma$	0179			
−2 05	$\mu - 2.05\sigma$	0202			
−2 00	$\mu - 2\ 00\sigma$	.0228	.05	$\mu + 05\sigma$	.5199
−1 95	$\mu - 1\ 95\sigma$	.0256	.10	$\mu + .10\sigma$	5398
−1.90	$\mu - 1\ 90\sigma$	0287	.15	$\mu + .15\sigma$	.5596
−1.85	$\mu - 1\ 85\sigma$	0322	.20	$\mu + .20\sigma$	.5793
−1 80	$\mu - 1\ 80\sigma$	.0359	.25	$\mu + .25\sigma$	.5987
−1 75	$\mu - 1\ 75\sigma$	0401	.30	$\mu + .30\sigma$	.6179
−1 70	$\mu - 1\ 70\sigma$	0446	.35	$\mu + .35\sigma$	.6368
−1 65	$\mu - 1\ 65\sigma$	0495	.40	$\mu + .40\sigma$	.6554
−1 60	$\mu - 1\ 60\sigma$	.0548	.45	$\mu + .45\sigma$	.6736
−1.55	$\mu - 1.55\sigma$	0606	.50	$\mu + .50\sigma$	.6915
−1.50	$\mu - 1\ 50\sigma$	0668	.55	$\mu + .55\sigma$	.7088
−1.45	$\mu - 1.45\sigma$	0735	.60	$\mu + .60\sigma$	.7257
−1 40	$\mu - 1\ 40\sigma$	0808	.65	$\mu + .65\sigma$	.7422
−1 35	$\mu - 1\ 35\sigma$	0885	.70	$\mu + .70\sigma$	.7580
−1.30	$\mu - 1\ 30\sigma$	0968	.75	$\mu + .75\sigma$	.7734
−1.25	$\mu - 1.25\sigma$	.1056	.80	$\mu + .80\sigma$	.7881
−1.20	$\mu - 1.20\sigma$	.1151	.85	$\mu + .85\sigma$	.8023
−1.15	$\mu - 1.15\sigma$	.1251	.90	$\mu + .90\sigma$	.8159
−1.10	$\mu - 1.10\sigma$	.1357	.95	$\mu + 95\sigma$	.8289
−1 05	$\mu - 1\ 05\sigma$	.1469	1.00	$\mu + 1.00\sigma$	.8413

Table A-5. Table of random numbers.

10	09	73	25	33	76	52	01	35	86	34	67	35	48	76	80	95	90	91	17	39	29	27	49	45
37	54	20	48	05	64	89	47	42	96	24	80	52	40	37	20	63	61	04	02	00	82	29	16	65
08	42	26	89	53	19	64	50	93	03	23	20	90	25	60	15	95	33	47	64	35	08	03	36	06
99	01	90	25	29	09	37	67	07	15	38	31	13	11	65	88	67	67	43	97	04	43	62	76	59
12	80	79	99	70	80	15	73	61	47	64	03	23	66	53	98	95	11	68	77	12	17	17	68	33
66	06	57	47	17	34	07	27	68	50	36	69	73	61	70	65	81	33	98	85	11	19	92	91	70
31	06	01	08	05	45	57	18	24	06	35	30	34	26	14	86	79	90	74	39	23	40	30	97	32
85	26	97	76	02	02	05	16	56	92	68	66	57	48	18	73	05	38	52	47	18	62	38	85	79
63	57	33	21	35	05	32	54	70	48	90	55	35	75	48	28	46	82	87	09	83	49	12	56	24
73	79	64	57	53	03	52	96	47	78	35	80	83	42	82	60	93	52	03	44	35	27	38	84	35
98	52	01	77	67	14	90	56	86	07	22	10	94	05	58	60	97	09	34	33	50	50	07	39	98
11	80	50	54	31	39	80	82	77	32	50	72	56	82	48	29	40	52	42	01	52	77	56	78	51
83	45	29	96	34	06	28	89	80	83	13	74	67	00	78	18	47	54	06	10	68	71	17	78	17
88	68	54	02	00	86	50	75	84	01	36	76	66	79	51	90	36	47	64	93	29	60	91	10	62
99	59	46	73	48	87	51	76	49	69	91	82	60	89	28	93	78	56	13	68	23	47	83	41	13
65	48	11	76	74	17	46	85	09	50	58	04	77	69	74	73	03	95	71	86	40	21	81	65	44
80	12	43	56	35	17	72	70	80	15	45	31	82	23	74	21	11	57	82	53	14	38	55	37	63
74	35	09	98	17	77	40	27	72	14	43	23	60	02	10	45	52	16	42	37	96	28	60	26	55
69	91	62	68	03	66	25	22	91	48	36	93	68	72	03	76	62	11	39	90	94	40	05	64	18
09	89	32	05	05	14	22	56	85	14	46	42	75	67	88	96	29	77	88	22	54	38	21	45	98
91	49	91	45	23	68	47	92	76	86	46	16	28	35	54	94	75	08	99	23	37	08	92	00	48
80	33	69	45	98	26	94	03	68	58	70	29	73	41	35	53	14	03	33	40	42	05	08	23	41
44	10	48	19	49	85	15	74	79	54	32	97	92	65	75	57	60	04	08	81	22	22	20	64	13
12	55	07	37	42	11	10	00	20	40	12	86	07	46	97	96	64	48	94	39	28	70	72	58	15
63	60	64	93	29	16	50	53	44	84	40	21	95	25	63	43	65	17	70	82	07	20	73	17	90
61	19	69	04	46	26	45	74	77	74	51	92	43	37	29	65	39	45	95	93	42	58	26	05	27
15	47	44	52	66	95	27	07	99	53	59	36	78	38	48	82	39	61	01	18	33	21	15	94	66
94	55	72	85	73	67	89	75	43	87	54	62	24	44	31	91	19	04	25	92	92	92	74	59	73
42	48	11	62	13	97	34	40	87	21	16	86	84	87	67	03	07	11	20	59	25	70	14	66	70
23	52	37	83	17	73	20	88	98	37	68	93	59	14	16	26	25	22	96	63	05	52	28	25	62
04	49	35	24	94	75	24	63	38	24	45	86	25	10	25	61	96	27	93	35	65	33	71	24	72
00	54	99	76	54	64	05	18	81	59	96	11	96	38	96	54	69	28	23	91	23	28	72	95	29
35	96	31	53	07	26	89	80	93	54	33	35	13	54	62	77	97	45	00	24	90	10	33	93	33
59	80	80	83	91	45	42	72	68	42	83	60	94	97	00	13	02	12	48	92	78	56	52	01	06
46	05	88	52	36	01	39	09	22	86	77	28	14	40	77	93	91	08	36	47	70	61	74	29	41
32	17	90	05	97	87	37	92	52	41	05	56	70	70	07	86	74	31	71	57	85	39	41	18	38
69	23	46	14	06	20	11	74	52	04	15	95	66	00	00	18	74	39	24	23	97	11	89	63	38
19	56	54	14	30	01	75	87	53	79	40	41	92	15	85	66	67	43	68	06	84	96	28	52	07
45	15	51	49	38	19	47	60	72	46	43	66	79	45	43	59	04	79	00	33	20	82	66	95	41
94	86	43	19	94	36	16	81	08	51	34	88	88	15	53	01	54	03	54	56	05	01	45	11	76
98	08	62	48	26	45	24	02	84	04	44	99	90	88	96	39	09	47	34	07	35	44	13	18	80
33	18	51	62	32	41	94	15	09	49	89	43	54	85	81	88	69	54	19	94	37	54	87	30	43
80	95	10	04	06	96	38	27	07	74	20	15	12	33	87	25	01	62	52	98	94	62	46	11	71
79	75	24	91	40	71	96	12	82	96	69	86	10	25	91	74	85	22	05	39	00	38	75	95	79
18	63	33	25	37	98	14	50	65	71	31	01	02	46	74	05	45	56	14	27	77	93	89	19	36
74	02	94	39	02	77	55	73	22	70	97	79	01	71	19	52	52	75	80	21	80	81	45	17	48
54	17	84	56	11	80	99	33	71	43	05	33	51	29	69	56	12	71	92	55	36	04	09	03	24
11	66	44	98	83	52	07	98	48	27	59	38	17	15	39	09	97	33	34	40	88	46	12	33	56
48	32	47	79	28	31	24	96	47	10	02	29	53	68	70	32	30	75	75	46	15	02	00	99	94
69	07	49	41	38	87	63	79	19	76	35	58	40	44	01	10	51	82	16	15	01	84	87	69	38

From tables of the RAND Corporation, by permission.

Table A-5. Table of random numbers (*Continued*).

09	18	82	00	97	32	82	53	95	27	04	22	08	63	04	83	38	98	73	74	64	27	85	80	44
90	04	58	54	97	51	98	15	06	54	94	93	88	19	97	91	87	07	61	50	68	47	66	46	59
73	18	95	02	07	47	67	72	62	69	62	29	06	44	64	27	12	46	70	18	41	36	18	27	60
75	76	87	64	90	20	97	18	17	49	90	42	91	22	72	95	37	50	58	71	93	82	34	31	78
54	01	64	40	56	66	28	13	10	03	00	68	22	73	98	20	71	45	32	95	07	70	61	78	13
08	35	86	99	10	78	54	24	27	85	13	66	15	88	73	04	61	89	75	53	31	22	30	84	20
28	30	60	32	64	81	33	31	05	91	40	51	00	78	93	32	60	46	04	75	94	11	90	18	40
53	84	08	62	33	81	59	41	36	28	51	21	59	02	90	28	46	66	87	95	77	76	22	07	91
91	75	75	37	41	61	61	36	22	69	50	26	39	02	12	55	78	17	65	14	83	48	34	70	55
89	41	59	26	94	00	39	75	83	91	12	60	71	76	46	48	94	97	23	06	94	54	13	74	08
77	51	30	38	20	86	83	42	99	01	68	41	48	27	74	51	90	81	39	80	72	89	35	55	07
19	50	23	71	74	69	97	92	02	88	55	21	02	97	73	74	28	77	52	51	65	34	46	74	15
21	81	85	93	13	93	27	88	17	57	05	68	67	31	56	07	08	28	50	46	31	85	33	84	52
51	47	46	64	99	68	10	72	36	21	94	04	99	13	45	42	83	60	91	91	08	00	74	54	49
99	55	96	83	31	62	53	52	41	70	69	77	71	28	30	74	81	97	81	42	43	86	07	28	34
33	71	34	80	07	93	58	47	28	69	51	92	66	47	21	58	30	32	98	22	93	17	49	39	72
85	27	48	68	93	11	30	32	92	70	28	83	43	41	37	73	51	59	04	00	71	14	84	36	43
84	13	38	96	40	44	03	55	21	66	73	85	27	00	91	61	22	26	05	61	62	32	71	84	23
56	73	21	62	34	17	39	59	61	31	10	12	39	16	22	85	49	65	75	60	81	60	41	88	80
65	13	85	68	06	87	64	88	52	61	34	31	36	58	61	45	87	52	10	69	85	64	44	72	77
38	00	10	21	76	81	71	91	17	11	71	60	29	29	37	74	21	96	40	49	65	58	44	96	98
37	40	29	63	97	01	30	47	75	86	56	27	11	00	86	47	32	46	26	05	40	03	03	74	38
97	12	54	03	48	87	08	33	14	17	21	81	53	92	50	75	23	76	20	47	15	50	12	95	78
21	82	64	11	34	47	14	33	40	72	64	63	88	59	02	49	13	90	64	41	03	85	65	45	52
73	13	54	27	42	95	71	90	90	35	85	79	47	42	96	08	78	98	81	56	64	69	11	92	02
07	63	87	79	29	03	06	11	80	72	96	20	74	41	56	23	82	19	95	38	04	71	36	69	94
60	52	88	34	41	07	95	41	98	14	59	17	52	06	95	05	53	35	21	39	61	21	20	64	55
83	59	63	56	55	06	95	89	29	83	05	12	80	97	19	77	43	35	37	83	92	30	15	04	98
10	85	06	27	46	99	59	91	05	07	13	49	90	63	19	53	07	57	18	39	06	41	01	93	62
39	82	09	89	52	43	62	26	31	47	64	42	18	08	14	43	80	00	93	51	31	02	47	31	67
59	58	00	64	78	75	56	97	88	00	88	83	55	44	86	23	76	80	61	56	04	11	10	84	08
38	50	80	73	41	23	79	34	87	63	90	82	29	70	22	17	71	90	42	07	95	95	44	99	53
30	69	27	06	68	94	68	81	61	27	56	19	68	00	91	82	06	76	34	00	05	46	26	92	00
65	44	39	56	59	18	28	82	74	37	49	63	22	40	41	08	33	76	56	76	96	29	99	08	36
27	26	75	02	64	13	19	27	22	94	07	47	74	46	06	17	98	54	89	11	97	34	13	03	58
91	30	70	69	91	19	07	22	42	10	36	69	95	37	28	28	82	53	57	93	28	97	66	62	52
68	43	49	46	88	84	47	31	36	22	62	12	69	84	08	12	84	38	25	90	09	81	59	31	46
48	90	81	58	77	54	74	52	45	91	35	70	00	47	54	83	82	45	26	92	54	13	05	51	60
06	91	34	51	97	42	67	27	86	01	11	88	30	95	28	63	01	19	89	01	14	97	44	03	44
10	45	51	60	19	14	21	03	37	12	91	34	23	78	21	88	32	58	08	51	43	66	77	08	83
12	88	39	73	43	65	02	76	11	84	04	28	50	13	92	17	97	41	50	77	90	71	22	67	69
21	77	83	09	76	38	80	73	69	61	31	64	94	20	96	63	28	10	20	23	08	81	64	74	49
19	52	35	95	15	65	12	25	96	59	86	28	36	82	58	69	57	21	37	98	16	43	59	15	29
67	24	55	26	70	35	58	31	65	63	79	24	68	66	86	76	46	33	42	22	26	65	59	08	02
60	58	44	73	77	07	50	03	79	92	45	13	42	65	29	26	76	08	36	37	41	32	64	43	44
53	85	34	13	77	36	06	69	48	50	58	83	87	38	59	49	36	47	33	31	96	24	04	36	42
24	63	73	87	36	74	38	48	93	42	52	62	30	79	92	12	36	91	86	01	03	74	28	38	73
83	08	01	24	51	38	99	22	28	15	07	75	95	17	77	97	37	72	75	85	51	97	23	78	67
16	44	42	43	34	36	15	19	90	73	27	49	37	09	39	85	13	03	25	52	54	84	65	47	59
60	79	01	81	57	57	17	86	57	62	11	16	17	85	76	45	81	95	29	79	65	13	00	48	60

Table A-6. Table of critical values for testing significance of outliers for T (one-sided test) when standard deviation is calculated from the complete sample.

Number of observations n	5% *Significance level*	2.5% *Significance level*	1% *Significance level*
3	1.15	1.15	1.15
4	1.46	1.48	1.49
5	1.67	1.71	1.75
6	1.82	1.89	1.94
7	1.94	2.02	2.10
8	2.03	2.13	2.22
9	2.11	2.21	2.32
10	2.18	2.29	2.41
11	2.23	2.36	2.48
12	2.29	2.41	2.55
13	2.33	2.46	2.61
14	2.37	2.51	2.66
15	2.41	2.55	2.71
16	2.44	2.59	2.75
17	2.47	2.62	2.79
18	2.50	2.65	2.82
19	2.53	2.68	2.85
20	2.56	2.71	2.88
21	2.58	2.73	2.91
22	2.60	2.76	2.94

Note. Values of T for $n \leqslant 25$ are based on those given in Reference [8]. For $n > 25$, the values of T are approximated. All values have been adjusted for division by $n - 1$ instead of n in calculating s.

By permission from *Technometrics* and Dr. Frank E. Grubbs, in *Technometrics*, February 1969.

Table A-6. Table of critical values for testing significance of outliers for T (one-sided test) when standard deviation is calculated from the complete sample (*Continued*).

Number of observations n	5% *Significance level*	2 5% *Significance level*	1% *Significance level*
23	2.62	2 78	2.96
24	2.64	2.80	2.99
25	2.66	2.82	3.01
30	2 75	2.91	
35	2.82	2.98	
40	2.87	3.04	
45	2 92	3 09	
50	2 96	3 13	
60	3 03	3 20	
70	3 09	3 26	
80	3.14	3 31	
90	3 18	3.35	
100	3 21	3 38	

$$T_n = \frac{x_n - \bar{x}}{s} \qquad s = \left\{\frac{\sum (x_i - \bar{x})^2}{n-1}\right\}^{\frac{1}{2}}$$

$$= \left\{\frac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)}\right\}^{\frac{1}{2}}$$

$$T_1 = \frac{\bar{x} - x_1}{s} \qquad x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$$

Table A-7. Percentiles of the chi-square distribution.

df	$P_{0.5}$	P_{01}	$P_{02.5}$	P_{05}	P_{10}	P_{90}	P_{95}	$P_{97.5}$	P_{99}	$P_{99.5}$
1	.000039	.00016	00098	.0039	0158	2 71	3 84	5 02	6 63	7 88
2	.0100	.0201	0506	1026	2107	4 61	5 99	7 38	9 21	10 60
3	.0717	115	.216	352	584	6 25	7 81	9 35	11 34	12 84
4	.207	297	484	711	1 064	7 78	9 49	11 14	13 28	14 86
5	.412	.554	.831	1.15	1 61	9 24	11 07	12 83	15 09	16 75
6	.676	.872	1 24	1.64	2 20	10 64	12 59	14 45	16 81	18 55
7	.989	1 24	1 69	2 17	2 83	12 02	14 07	16 01	18 48	20 28
8	1 34	1 65	2 18	2 73	3 49	13 36	15.51	17 53	20 09	21 96
9	1 73	2 09	2.70	3.33	4.17	14 68	16 92	19 02	21 67	23 59
10	2 16	2 56	3.25	3.94	4 87	15 99	18 31	20 48	23 21	25 19
11	2 60	3 05	3.82	4 57	5 58	17 28	19 68	21 92	24.73	26 76
12	3.07	3 57	4.40	5 23	6 30	18 55	21 03	23 34	26 22	28 30
13	3 57	4 11	5 01	5 89	7 04	19.81	22 36	24 74	27 69	29 82
14	4.07	4 66	5.63	6 57	7 79	21.06	23 68	26 12	29.14	31 32
15	4.60	5 23	6 26	7 26	8 55	22 31	25 00	27 49	30.58	32 80
16	5 14	5.81	6 91	7 96	9 31	23 54	26 30	28 85	32 00	34.27
18	6 26	7.01	8.23	9 39	10 86	25 99	28 87	31.53	34 81	37 16
20	7.43	8.26	9.59	10 85	12 44	28 41	31 41	34 17	37 57	40 00
24	9.89	10 86	12.40	13 85	15 66	33 20	36 42	39.36	42 98	45 56
30	13.79	14.95	16.79	18 49	20.60	40.26	43.77	46 98	50 89	53 67
40	20 71	22 16	24.43	26 51	29 05	51.81	55 76	59 34	63 69	66 77
60	35.53	37.48	40.48	43 19	46 46	74.40	79 08	83.30	88 38	91.95
120	83.85	86.92	91.58	95 70	100.62	140 23	146 57	152 21	158 95	163 64

Table A-8. Percentiles of student's *t* distribution.

df	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	.325	.727	1 376	3 078	6 314	12 706	31 821	63 657
2	.289	617	1.061	1 886	2 920	4 303	6 965	9 925
3	.277	.584	.978	1 638	2 353	3 182	4.541	5 841
4	.271	.569	.941	1 533	2 132	2 776	3 747	4.604
5	.267	.559	.920	1.476	2 015	2 571	3 365	4.032
6	.265	.553	.906	1.440	1.943	2.447	3 143	3 707
7	.263	.549	.896	1 415	1.895	2.365	2 998	3.499
8	.262	.546	.889	1.397	1 860	2 306	2 896	3.355
9	.261	.543	.883	1.383	1 833	2 262	2 821	3.250
10	.260	.542	.879	1 372	1 812	2 228	2 764	3.169
11	.260	.540	.876	1.363	1 796	2.201	2.718	3.106
12	.259	.539	.873	1 356	1 782	2.179	2 681	3 055
13	.259	.538	.870	1 350	1.771	2.160	2 650	3.012
14	.258	.537	868	1 345	1 761	2.145	2 624	2.977
15	.258	.536	.866	1 341	1 753	2.131	2 602	2.947
16	.258	.535	.865	1 337	1 746	2.120	2 583	2.921
17	.257	.534	.863	1 333	1 740	2.110	2 567	2.898
18	257	.534	.862	1 330	1 734	2.101	2 552	2.878
19	.257	.533	.861	1 328	1.729	2 093	2.539	2.861
20	257	.533	.860	1 325	1.725	2.086	2 528	2.845
21	.257	.532	.859	1 323	1 721	2 080	2.518	2.831
22	.256	.532	.858	1 321	1 717	2.074	2 508	2 819
23	256	.532	.858	1 319	1.714	2 069	2 500	2 807
24	.256	.531	.857	1.318	1 711	2 064	2 492	2.797
25	.256	.531	.856	1.316	1 708	2 060	2.485	2.787
26	.256	.531	.856	1.315	1.706	2.056	2 479	2.779
27	256	531	.855	1.314	1.703	2 052	2 473	2.771
28	256	530	.855	1 313	1.701	2.048	2.467	2.763
29	.256	.530	.854	1.311	1.699	2.045	2.462	2.756
30	256	530	.854	1 310	1.697	2.042	2 457	2.750
40	.255	.529	.851	1.303	1.684	2 021	2.423	2 704
60	.254	.527	848	1.296	1.671	2.000	2.390	2.660
120	.254	.526	.845	1 289	1.658	1.980	2.358	2 617
∞	.253	.524	.842	1 282	1.645	1.960	2.326	2.576
df	$-t_{.40}$	$-t_{.30}$	$-t_{.20}$	$-t_{.10}$	$-t_{.05}$	$-t_{.025}$	$-t_{.01}$	$-t_{.005}$

Table A-9. Percentiles of the *F* distribution.

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

ν_2, DEGREES OF FREEDOM FOR DENOMINATOR	*Cum. prop.*	1	2	3	4	5	6	7	8	9	10	11	12	*Cum. prop.*
1	.0005	.0^{6}62	0^{3}50	.0^{2}38	0^{2}94	016	.022	027	032	036	039	042	045	0005
	001	.0^{5}25	0^{2}10	.0^{2}60	013	021	028	.034	039	044	048	051	.054	001
	.005	.0^{4}62	0^{2}51	.018	032	044	.054	.062	.068	.073	078	082	.085	005
	.010	.0^{3}25	010	.029	047	062	073	.082	.089	095	100	104	107	010
	.025	.0^{2}15	.026	.057	.082	.100	.113	.124	.132	139	144	149	.153	025
	.05	.0^{2}62	.054	.099	.130	.151	.167	.179	.188	195	201	207	.211	05
	.10	.025	117	.181	.220	.246	265	.279	.289	298	304	310	.315	10
	.25	.172	389	.494	553	.591	617	.637	.650	661	670	680	684	25
	.50	1 00	1.50	1.71	1 82	1.89	1.94	1 98	2 00	2 03	2 04	2 05	2 07	50
	.75	5 83	7 50	8 20	8 58	8.82	8 98	9 10	9 19	9 26	9 32	9 36	9.41	75
	.90	39 9	49 5	53.6	55 8	57.2	58 2	58 9	59 4	59 9	60 2	60.5	60 7	90
	.95	161	200	216	225	230	234	237	239	241	242	243	244	.95
	975	648	800	864	900	922	937	948	957	963	969	973	977	975
	.99	405^1	500^1	540^1	562^1	576^1	586^1	593^1	598^1	602^1	606^1	608^1	611^1	99
	.995	162^2	200^2	216^2	225^2	231^2	234^2	237^2	239^2	241^2	242^2	243^2	244^2	995
	.999	406^3	500^3	540^3	562^3	576^3	586^3	593^3	598^3	602^3	606^3	609^3	611^3	999
	.9995	162^4	200^4	216^4	225^4	231^4	234^4	237^4	239^4	241^4	242^4	243^4	244^4	9995
2	.0005	.0^{6}50	0^{3}50	.0^{2}42	.011	020	.029	037	.044	050	056	.061	065	0005
	.001	0^{5}20	0^{2}10	.0^{2}68	.016	027	037	.046	.054	.061	067	072	077	001
	.005	.0^{4}50	0^{2}50	.020	.038	055	069	.081	.091	099	106	112	118	005
	.01	0^{3}20	.010	032	056	075	092	.105	.116	.125	.132	.139	144	01
	.025	.0^{2}13	.026	062	094	.119	138	.153	.165	175	183	190	196	025
	.05	.0^{2}50	053	.105	144	.173	194	.211	.224	.235	244	251	.257	05
	.10	.020	.111	.183	231	.265	289	.307	.321	.333	342	350	.356	.10
	.25	133	.333	.439	500	540	.568	588	.604	.616	.626	633	641	.25
	.50	.667	1 00	1.13	1 21	1 25	1.28	1.30	1.32	1 33	1 34	1 35	1 36	.50
	.75	2 57	3 00	3.15	3.23	3 28	3.31	3 34	3 35	3 37	3 38	3 39	3 39	.75
	.90	8 53	9.00	9.16	9.24	9 29	9 33	9 35	9.37	9 38	9.39	9.40	9 41	90
	.95	18.5	19 0	19.2	19.2	19.3	19 3	19.4	19.4	19 4	19 4	19.4	19 4	.95
	.975	38 5	39 0	39.2	39 2	39.3	39.3	39.4	39 4	39 4	39 4	39.4	39.4	975
	.99	98 5	99 0	99 2	99.2	99.3	99 3	99 4	99 4	99 4	99 4	99 4	99 4	.99
	.995	198	199	199	199	199	199	199	199	199	199	199	199	995
	.999	998	999	999	999	999	999	999	999	999	999	999	999	.999
	.9995	200^1	200^1	200^1	200^1	200^1	200^1	200^1	200^1	200^1	200^1	200^1	200^1	.9995
3	.0005	0^{6}46	.0^{3}50	0^{2}44	.012	.023	.033	043	.052	.060	.067	074	079	0005
	001	.0^{5}19	.0^{2}10	0^{2}71	018	030	.042	053	.063	.072	.079	.086	.093	001
	.005	.0^{4}46	.0^{2}50	021	041	.060	077	.092	104	.115	124	132	138	.005
	.01	.0^{3}19	.010	034	060	083	102	118	132	.143	153	161	168	.01
	.025	.0^{2}12	.026	.065	.100	.129	.152	170	.185	.197	207	216	224	.025
	.05	.0^{2}46	.052	.108	152	.185	.210	230	.246	.259	.270	279	.287	.05
	.10	019	.109	.185	.239	276	.304	.325	342	.356	.367	.376	384	10
	.25	122	.317	.424	.489	.531	.561	582	600	.613	.624	633	641	.25
	.50	585	.881	1.00	1 06	1 10	1.13	1.15	1 16	1 17	1 18	1.19	1.20	50
	.75	2 02	2 28	2 36	2 39	2.41	2 42	2 43	2 44	2.44	2 44	2 45	2 45	75
	.90	5.54	5 46	5 39	5 34	5 31	5 28	5 27	5 25	5 24	5 23	5.22	5 22	90
	.95	10.1	9 55	9 28	9 12	9.01	8 94	8.89	8 85	8 81	8 79	8 76	8 74	95
	.075	17.4	16 0	15.4	15.1	14 9	14 7	14.6	14 5	14 5	14 4	14 4	14 3	.975
	.99	34.1	30 8	29.5	28 7	28 2	27.9	27 7	27 5	27 3	27 2	27 1	27 1	99
	.995	55.6	49 8	47.5	46 2	45 4	44.8	44.4	44 1	43 9	43 7	43.5	43.4	995
	.999	167	149	141	137	135	133	132	131	130	129	129	128	.999
	.9995	266	237	225	218	214	211	209	208	207	206	204	204	.9995

Read .0^{3}56 as .00056, 200^1 as 2,000, 162^4 as 1,620,000, and so on.

Table A-9. Percentiles of the F distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

Cum. prop.	15	20	24	30	40	50	60	100	120	200	500	∞	*Cum prop.*	ν_2, DEGREES OF FREEDOM FOR DENOMINATOR
0005	.051	058	062	066	069	072	.074	077	078	080	081	083	0005	1
001	.060	.067	071	075	079	082	084	087	088	089	091	092	001	
005	.093	.101	.105	.109	113	.116	118	.121	122	124	.126	127	005	
01	.115	124	.128	.132	.137	139	141	145	146	148	.150	151	01	
025	.161	170	175	.180	184	187	189	.193	194	196	.198	199	025	
05	.220	.230	235	.240	245	248	250	.254	255	257	.259	261	05	
.10	.325	336	342	347	353	356	358	.362	364	366	.368	370	10	
25	.698	712	.719	727	734	738	741	.747	749	752	.754	756	25	
50	2 09	2 12	2 13	2.15	2 16	2.17	2.17	2 18	2 18	2.19	2 19	2 20	50	
75	9 49	9 58	9 63	9.67	9 71	9 74	9 76	9 78	9 80	9 82	9 84	9 85	75	
90	61.2	61 7	62.0	62 3	62 5	62 7	62 8	63 0	63.1	63 2	63 3	63 3	90	
.95	246	248	249	250	251	252	252	253	253	254	254	254	.95	
.975	985	993	997	100^1	101^1	101^1	101^1	101^1	101^1	102^1	102^1	102^1	.975	
.99	616^1	621^1	623^1	626^1	629^1	630^1	631^1	633^1	634^1	635^1	636^1	637^1	99	
.995	246^2	248^2	249^2	250^2	251^2	252^2	253^2	253^2	254^2	254^2	254^2	255^2	995	
.999	616^3	621^3	623^3	626^3	629^3	630^3	631^3	633^3	634^3	635^3	636^3	637^3	999	
.9995	246^4	248^4	249^4	250^4	251^4	252^4	252^4	253^4	253^4	253^4	254^4	254^4	.9995	
.0005	.076	088	094	.101	.108	113	116	.122	124	127	.130	132	0005	2
.001	.088	100	.107	.114	.121	126	.129	.135	.137	140	.143	145	001	
.005	.130	143	150	157	.165	169	173	.179	.181	184	.187	189	005	
.01	.157	171	178	186	.193	198	201	.207	209	212	.215	217	01	
.025	.210	.224	.232	.239	247	251	255	.261	263	266	.269	.271	025	
.05	.272	286	.294	.302	309	314	317	.324	.326	329	.332	334	.05	
.10	.371	386	.394	.402	410	415	418	.424	426	429	.433	.434	.10	
.25	.657	.672	.680	.689	697	702	.705	.711	713	716	.719	721	.25	
50	1.38	1.39	1.40	1 41	1 42	1 42	1.43	1.43	1 43	1.44	1 44	1 44	50	
.75	3 41	3 43	3.43	3 44	3.45	3.45	3.46	3 47	3 47	3 48	3 48	3 48	75	
.90	9.42	9 44	9.45	9.46	9.47	9.47	9.47	9 48	9.48	9 49	9 49	9.49	90	
.95	19.4	19 4	19 5	19.5	19.5	19 5	19 5	19.5	19 5	19 5	19 5	19.5	.95	
.975	39 4	39.4	39 5	39 5	39.5	39 5	39.5	39.5	39 5	39 5	39.5	39.5	975	
.99	99.4	99.4	99 5	99.5	99.5	99 5	99.5	99.5	99 5	99 5	99 5	99 5	99	
.995	199	199	199	199	199	199	199	199	199	199	199	200	.995	
.999	999	999	999	999	999	999	999	999	999	999	999	999	.999	
9995	200^1	200^1	200^1	200^1	200^1	200^1	200^1	200^1	200^1	200^1	200^1	200^1	.9995	
.0005	.093	.109	.117	127	136	.143	.147	.156	158	162	.166	169	0005	3
001	.107	123	.132	.142	152	.158	.162	.171	.173	177	.181	.184	.001	
.005	.154	.172	.181	191	.201	207	.211	220	222	227	.231	.234	.005	
.01	.185	.203	.212	.222	232	.238	.242	.251	253	258	.262	264	.01	
.025	.241	259	.269	.279	.289	.295	.299	.308	310	314	318	.321	025	
.05	.304	.323	.332	.342	.352	358	.363	.370	373	377	.382	.384	.05	
.10	.402	420	.430	439	449	455	.459	467	.469	.474	.476	480	.10	
25	.658	.675	.684	.693	.702	708	.711	719	721	724	728	730	.25	
.50	1.21	1.23	1 23	1.24	1.25	1.25	1 25	1.26	1.26	1.26	1 27	1 27	.50	
.75	2.46	2.46	2.46	2.47	2.47	2.47	2 47	2.47	2 47	2 47	2.47	2.47	.75	
.90	5.20	5.18	5.18	5.17	5.16	5.15	5 15	5.14	5.14	5.14	5 14	5.13	.90	
.95	8.70	8 66	8.63	8.62	8.59	8 58	8 57	8.55	8.55	8.54	8 53	8 53	.95	
.975	14 3	14.2	14.1	14.1	14.0	14.0	14.0	14 0	13 9	13 9	13 9	13 9	.975	
.99	26 9	26.7	26.6	26.5	26 4	26 4	26.3	26.2	26 2	26.2	26.1	26 1	.99	
.995	43.1	42 8	42.6	42.5	42 3	42 2	42.1	42 0	42 0	41.9	41.9	41 8	.995	
.999	127	126	126	125	125	125	124	124	124	124	124	123	.999	
.9995	203	201	200	199	199	198	198	197	197	197	196	196	.9995	

Table A-9. Percentiles of the *F* distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

ν_2, DEGREES OF FREEDOM FOR DENOMINATOR	*Cum prop*	1	2	3	4	5	6	7	8	9	10	11	12	*Cum. prop.*
4	0005	$0^{6}44$	$0^{3}50$	$0^{2}46$	013	024	036	047	057	066	075	082	089	0005
	001	$0^{5}18$	$0^{2}10$	$0^{2}73$	019	032	046	058	069	079	089	097	104	001
	005	$0^{4}44$	$0^{2}50$	022	043	064	083	100	114	126	137	145	153	005
	01	$0^{3}18$	.010	035	063	088	109	127	143	156	167	176	185	01
	025	$0^{2}11$	.026	066	104	135	161	181	198	212	224	.234	243	025
	05	$0^{2}44$	.052	.110	157	.193	221	243	261	275	288	298	307	05
	10	018	108	187	243	284	.314	338	356	371	384	394	403	10
	25	117	.309	418	484	528	560	583	601	615	627	637	645	25
	50	549	828	941	1 00	1 04	1 06	1 08	1 09	1 10	1 11	1 12	1 13	50
	75	1.81	2 00	2 05	2 06	2 07	2 08	2 08	2 08	2 08	2 08	2 08	2 08	75
	90	4 54	4 32	4.19	4.11	4 05	4.01	3 98	3 95	3.94	3 92	3 91	3 90	90
	95	7 71	6 94	6 59	6.39	6 26	6 16	6 09	6 04	6 00	5 96	5 94	5 91	95
	.975	12 2	10 6	9 98	9 60	9 36	9.20	9 07	8 98	8 90	8 84	8 79	8 75	975
	99	21.2	18 0	16 7	16 0	15.5	15 2	15 0	14 8	14 7	14 5	14 4	14 4	99
	.995	31.3	26 3	24.3	23 2	22.5	22 0	21 6	21 4	21 1	21 0	20 8	20 7	995
	999	74.1	61 2	56.2	53 4	51.7	50 5	49 7	49 0	48 5	48 0	47 7	47.4	999
	.9995	106	87 4	80 1	76 1	73 6	71 9	70 6	69 7	68 9	68 3	67 8	67 4	9995
5	0005	$.0^{6}43$	$0^{3}50$	$0^{2}47$	014	025	038	050	061	070	081	089	.096	0005
	001	$.0^{5}17$	$0^{2}10$	$0^{2}75$	019	034	048	062	074	085	.095	104	.112	001
	005	$0^{4}43$	$0^{2}50$	022	045	067	087	.105	120	134	146	156	.165	005
	01	$0^{3}17$	010	035	064	091	114	.134	151	165	177	188	197	01
	025	$0^{2}11$	025	067	107	140	167	.189	.208	223	236	248	257	025
	.05	$.0^{2}43$	052	111	160	198	228	.252	271	287	301	313	322	05
	10	017	108	188	247	290	322	.347	367	383	397	408	418	10
	25	.113	305	415	483	528	560	584	604	618	631	641	650	25
	50	528	799	907	965	1.00	1 02	1.04	1 05	1 06	1 07	1.08	1 09	50
	75	1 69	1 85	1 88	1.89	1 89	1 89	1 89	1 89	1 89	1 89	1.89	1 89	75
	.90	4 06	3.78	3 62	3 52	3 45	3 40	3 37	3 34	3 32	3.30	3.28	3 27	90
	95	6 61	5 79	5 41	5.19	5.05	4 95	4 88	4 82	4.77	4.74	4.71	4 68	95
	975	10 0	8 43	7.76	7 39	7.15	6 98	6 85	6 76	6 68	6 62	6 57	6 52	975
	99	16 3	13 3	12 1	11 4	11 0	10 7	10 5	10 3	10 2	10 1	9.96	9 89	99
	995	22 8	18 3	16 5	15 6	14 9	14.5	14 2	14.0	13 8	13 6	13 5	13.4	995
	999	47 2	37 1	33 2	31 1	29 7	28.8	28 2	27 6	27 2	26 9	26 6	26 4	999
	9995	63.6	49.8	44 4	41.5	39 7	38.5	37 6	36 9	36 4	35 9	35 6	35 2	.9995
6	0005	$.0^{6}43$	$0^{3}50$	$.0^{2}47$	014	026	.039	.052	064	075	085	094	103	0005
	001	$.0^{5}17$	$.0^{2}10$	$0^{2}75$	020	035	050	064	.078	090	101	111	119	001
	005	$0^{4}43$	$0^{2}50$	022	045	069	090	109	126	140	153	164	174	005
	01	$0^{3}17$	010	.036	066	094	118	.139	.157	172	186	197	207	01
	025	$.0^{2}11$	.025	.068	109	143	172	.195	.215	.231	246	258	268	025
	05	$0^{2}43$	.052	112	162	202	233	.259	279	.296	311	324	334	05
	10	017	.107	.189	249	294	.327	.354	375	392	406	418	429	.10
	25	111	.302	413	481	524	561	.586	606	622	635	645	654	25
	50	515	780	.886	942	977	1.00	1 02	1.03	1.04	1 05	1 05	1 06	50
	75	1 62	1 76	1 78	1 79	1 79	1.78	1 78	1 78	1 77	1.77	1.77	1 77	.75
	90	3 78	3 46	3.29	3 18	3 11	3 05	3 01	2 98	2.96	2.94	2.92	2 90	90
	95	5 99	5 14	4.76	4 53	4 39	4 28	4.21	4 15	4 10	4 06	4 03	4 00	95
	975	8 81	7 26	6.60	6 23	5 99	5 82	5 70	5 60	5 52	5 46	5.41	5.37	975
	99	13 7	10 9	9 78	9 15	8 75	8 47	8 26	8 10	7 98	7 87	7 79	7.72	.99
	995	18 6	14 5	12 9	12 0	11 5	11 1	10 8	10.6	10 4	10 2	10 1	10.0	.995
	999	35 5	27 0	23.7	21 9	20 8	20 0	19.5	19 0	18 7	18 4	18 2	18.0	999
	9995	46 1	34.8	30 4	28 1	26 6	25 6	24.9	24 3	23 9	23.5	23.2	23 0	.9995

Table A-9. Percentiles of the *F* distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

Cum. prop.	15	20	24	30	40	50	60	100	120	200	500	∞	*Cum prop.*	ν_2, DEGREES OF FREEDOM FOR DENOMINATOR
0005	105	125	135	147	159	166	172	183	186	191	196	200	0005	4
001	121	141	152	163	176	183	.188	200	202	208	213	217	001	
005	172	193	204	216	.229	237	242	253	255	260	266	269	005	
01	204	226	237	249	.261	269	274	285	287	293	298	301	01	
025	263	284	.296	308	.320	.327	.332	342	.346	351	356	359	025	
05	.327	349	360	372	384	.391	.396	407	409	.413	418	422	.05	
10	424	.445	456	467	.478	.485	.490	500	.502	.508	510	514	10	
.25	664	.683	692	.702	.712	.718	.722	731	733	737	740	743	25	
50	1.14	1 15	1 16	1 16	1 17	1 18	1.18	1 18	1.18	1 19	1 19	1 19	50	
.75	2 08	2.08	2 08	2 08	2 08	2 08	2 08	2.08	2 08	2 08	2 08	2 08	75	
90	3.87	3.84	3 83	3 82	3 80	3.80	3 79	3 78	3 78	3.77	3 76	3 76	90	
95	5.86	5 80	5 77	5 75	5 72	5 70	5.69	5 66	5 66	5 65	5 64	5 63	95	
975	8 66	8 56	8.51	8 46	8 41	8 38	8 36	8 32	8 31	8 29	8 27	8 26	975	
99	14.2	14 0	13 9	13.8	13 7	13.7	13 7	13 6	13 6	13 5	13 5	13 5	99	
995	20 4	20.2	20.0	19.9	19 8	19 7	19.6	19 5	19 5	19 4	19 4	19 3	995	
999	46 8	46.1	45 8	45.4	45 1	44 9	44.7	44 5	44 4	44.3	44.1	44 0	999	
9995	66.5	65 5	65.1	64.6	64 1	63 8	63 6	63 2	63 1	62.9	62 7	62 6	9995	
0005	.115	.137	150	163	177	.186	192	205	209	.216	222	.226	0005	5
001	132	155	167	.181	195	.204	210	223	227	.233	239	244	001	
005	.186	210	.223	237	.251	260	266	279	282	.288	294	299	.005	
01	219	244	257	270	.285	.293	299	312	315	322	328	.331	01	
025	280	304	.317	.330	344	353	359	370	374	.380	386	390	025	
.05	.345	.369	.382	395	.408	.417	422	432	.437	.442	448	.452	05	
.10	.440	463	.476	.488	.501	508	514	524	.527	532	538	541	10	
.25	669	690	700	.711	722	728	732	741	.743	.748	752	755	25	
.50	1.10	1 11	1.12	1 12	1 13	1 13	1 14	1.14	1 14	1 15	1.15	1.15	.50	
.75	1.89	1 88	1.88	1 88	1 88	1.88	1.87	1.87	1 87	1 87	1 87	1.87	.75	
.90	3.24	3.21	3 19	3.17	3.16	3 15	3.14	3 13	3 12	3 12	3 11	3.10	90	
.95	4 62	4.56	4 53	4.50	4 46	4 44	4 43	4 41	4 40	4 39	4.37	4 36	.95	
975	6 43	6.33	6 28	6 23	6.18	6.14	6.12	6 08	6 07	6 05	6 03	6 02	975	
.99	9 72	9.55	9.47	9 38	9.29	9.24	9 20	9 13	9 11	9 08	9.04	9.02	99	
995	13 1	12 9	12.8	12.7	12 5	12 5	12 4	12 3	12 3	12.2	12 2	12 1	995	
999	25.9	25.4	25.1	24.9	24.6	24.4	24.3	24 1	24 1	23.9	23 8	23 8	.999	
.9995	34 6	33 9	33.5	33.1	32 7	32 5	32.3	32.1	32.0	31.8	31 7	31.6	.9995	
.0005	.123	.148	.162	.177	.193	203	210	225	229	.236	.244	.249	.0005	6
.001	.141	.166	180	.195	.211	222	.229	.243	247	.255	262	.267	.001	
.005	.197	.224	.238	253	.269	.279	286	.301	.304	.312	318	.324	.005	
.01	.232	.258	.273	.288	.304	.313	321	.334	.338	346	352	.357	.01	
.025	.293	.320	.334	.349	.364	.375	381	.394	.398	405	412	.415	.025	
.05	.358	.385	.399	.413	.428	.437	444	.457	.460	.467	.472	.476	.05	
.10	.453	.478	.491	.505	.519	526	533	.546	.548	556	559	564	.10	
.25	.675	.696	.707	.718	.729	.736	.741	.751	.753	.758	762	.765	.25	
.50	1 07	1.08	1 09	1.10	1 10	1 11	1.11	1.11	1 12	1 12	1.12	1.12	50	
.75	1.76	1 76	1 75	1.75	1.75	1 75	1.74	1.74	1.74	1 74	1.74	1 74	.75	
.90	2 87	2.84	2.82	2.80	2.78	2.77	2.76	2 75	2 74	2.73	2.73	2.72	.90	
.95	3 94	3.87	3 84	3 81	3.77	3.75	3.74	3 71	3 70	3 69	3.68	3.67	95	
.975	5.27	5 17	5 12	5 07	5 01	4 98	4.96	4 92	4.90	4.88	4.86	4.85	.975	
.99	7 56	7 40	7.31	7.23	7.14	7 09	7 06	6 99	6 97	6.93	6.90	6.88	.99	
.995	9 81	9 59	9 47	9 36	9 24	9 17	9 12	9 03	9 00	8 95	8.91	8.88	.995	
.999	17 6	17.1	16.9	16 7	16 4	16 3	16 2	16.0	16 0	15.9	15.8	15 7	.999	
.9995	22 4	21.9	21.7	21.4	21.1	20 9	20 7	20.5	20 4	20.3	20 2	20 1	.9995	

Table A-9. Percentiles of the *F* distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

ν_2, DEGREES OF FREEDOM FOR DENOMINATOR	*Cum. prop.*	1	2	3	4	5	6	7	8	9	10	11	12	*Cum prop.*
7	0005	$0^{6}42$	$0^{3}50$	$0^{2}48$	014	.027	040	053	066	078	088	099	108	0005
	001	$0^{5}17$	$0^{2}10$	$0^{2}76$	020	.035	051	067	081	093	105	115	125	001
	005	$0^{4}42$	$0^{2}50$	023	.046	070	093	113	130	145	159	171	181	005
	01	$.0^{3}17$	010	036	067	096	121	143	162	178	192	205	216	01
	025	$0^{2}10$	025	068	.110	146	176	200	221	.238	253	266	277	025
	05	$0^{2}42$	052	113	164	.205	238	264	286	.304	319	332	.343	05
	10	017	107	.190	.251	297	.332	359	381	.399	414	427	438	10
	25	110	300	.412	.481	528	562	588	608	624	637	649	658	25
	50	506	767	871	926	960	983	1 00	1 01	1 02	1 03	1 04	1.04	50
	.75	1 57	1 70	1 72	1 72	1 71	1 71	1 70	1 70	1 69	1 69	1 69	1 68	75
	90	3 59	3 26	3 07	2 96	2 88	2 83	2 78	2 75	2 72	2 70	2 68	2 67	90
	95	5 59	4.74	4 35	4.12	3 97	3 87	3 79	3 73	3 68	3 64	3 60	3 57	95
	.975	8 07	6 54	5 89	5 52	5 29	5 12	4 99	4 90	4 82	4 76	4 71	4 67	975
	99	12 2	9 55	8 45	7 85	7 46	7 19	6 99	6 84	6 72	6 62	6 54	6 47	99
	995	16 2	12 4	10 9	10 0	9 52	9.16	8 89	8 68	8 51	8 38	8 27	8 18	995
	.999	29 2	21 7	18 8	17 2	16 2	15 5	15 0	14 6	14 3	14.1	13 9	13 7	999
	.9995	37 0	27 2	23 5	21 4	20 2	19 3	18 7	18 2	17 8	17 5	17 2	17 0	9995
8	.0005	$0^{6}42$	$0^{3}50$	$0^{2}48$	014	027	041	055	068	081	.092	102	112	0005
	001	$0^{5}17$	$0^{2}10$	$0^{2}76$	020	036	053	068	083	.096	.109	120	130	001
	005	$0^{4}42$	$0^{2}50$	027	047	072	095	115	133	149	.164	176	187	005
	01	$0^{3}17$	010	036	068	097	123	146	166	.183	198	211	222	01
	025	$0^{2}10$	025	069	111	148	179	204	226	244	.259	273	285	025
	05	$0^{2}42$	052	.113	166	.208	241	268	291	310	.326	339	351	05
	10	017	107	190	253	299	335	363	386	405	.421	435	445	10
	.25	109	298	411	481	529	563	589	610	627	640	654	661	25
	50	499	757	860	915	948	971	988	1 00	1 01	1 02	1 02	1 03	50
	75	1 54	1 66	1.67	1 66	1 66	1 65	1 64	1 64	1 64	1 63	1 63	1 62	75
	90	3 46	3 11	2.92	2 81	2 73	2 67	2.62	2 59	2 56	2 54	2 52	2 50	90
	95	5 32	4 46	4 07	3 84	3 69	3 58	3 50	3 44	3 39	3 35	3 31	3 28	95
	.975	7 57	6 06	5 42	5 05	4 82	4 65	4.53	4 43	4 36	4 30	4 24	4 20	.975
	99	11 3	8 65	7 59	7 01	6 63	6 37	6.18	6 03	5.91	5 81	5 73	5 67	99
	995	14 7	11 0	9 60	8 81	8 30	7 95	7 69	7 50	7 34	7 21	7 10	7 01	995
	.999	25 4	18 5	15 8	14.4	13 5	12 9	12 4	12 0	11 8	11 5	11.4	11 2	999
	9995	31 6	22 8	19 4	17 6	16 4	15 7	15 1	14 6	14 3	14.0	13.8	13 6	9995
9	0005	$0^{6}41$	$0^{3}50$	$0^{2}48$	015	027	042	056	070	083	094	105	115	0005
	001	$0^{5}17$	$0^{2}10$	$0^{2}77$	021	037	054	070	085	099	112	123	134	001
	005	$0^{4}42$	$0^{2}50$	023	047	073	096	117	136	153	168	181	192	005
	01	$.0^{3}17$	010	037	068	098	125	149	169	187	202	216	228	01
	025	$0^{2}10$	025	069	112	150	181	207	230	248	265	279	291	025
	05	$0^{2}40$	052	113	167	210	244	272	296	315	331	345	358	05
	10	017	107	191	254	302	.338	367	390	410	426	441	452	10
	25	.108	297	410	480	529	564	591	612	629	643	654	664	25
	50	494	749	852	906	939	962	978	990	1 00	1 01	1 01	1 02	50
	75	1 51	1 62	1 63	1 63	1 62	1 61	1 60	1 60	1 59	1 59	1 58	1 58	75
	.90	3 36	3 01	2 81	2 69	2 61	2 55	2 51	2.47	2 44	2 42	2 40	2 38	90
	95	5 12	4 26	3 86	3 63	3 48	3 37	3 29	3 23	3 18	3 14	3 10	3 07	95
	.975	7 21	5 71	5 08	4 72	4 48	4 32	4 20	4 10	4 03	3 96	3 91	3 87	975
	.99	10 6	8 02	6 99	6 42	6 06	5 80	5 61	5 47	5 35	5 26	5 18	5 11	99
	.995	13 6	10 1	8 72	7 96	7 47	7 13	6 88	6 69	6 54	6 42	6 31	6 23	995
	.999	22 9	16 4	13 9	12 6	11 7	11 1	10 7	10 4	10 1	9 89	9 71	9 57	999
	.9995	28 0	19 9	16 8	15 1	14 1	13 3	12 8	12 4	12 1	11 8	11 6	11 4	9995

Table A-9. Percentiles of the *F* distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

Cum. prop.	15	20	24	30	40	50	60	100	120	200	500	∞	*Cum. prop.*	ν_2, DEGREES OF FREEDOM FOR DENOMINATOR
.0005	130	157	172	188	206	217	.225	242	246	255	263	268	0005	7
.001	148	176	191	208	225	237	.245	261	266	274	282	288	001	
.005	206	235	251	267	285	296	304	319	324	332	340	345	005	
.01	241	.270	286	303	320	331	339	355	358	366	373	379	01	
025	304	333	348	.364	381	392	399	413	418	426	433	437	025	
.05	369	398	413	428	445	455	.461	476	479	485	493	498	05	
.10	.463	491	504	519	534	543	550	562	566	571	578	582	10	
.25	679	702	713	725	.737	745	749	760	762	767	772	775	25	
.50	1 05	1 07	1 07	1 08	1 08	1 09	1 09	1 10	1 10	1 10	1 10	1 10	50	
.75	1 68	1 67	1 67	1 66	1 66	1 66	1 65	1 65	1 65	1 65	1 65	1 65	75	
.90	2 63	2 59	2 58	2 56	2 54	2 52	2 51	2 50	2 49	2 48	2 48	2 47	90	
.95	3 51	3 44	3 41	3 38	3 34	3 32	3 30	3 27	3 27	3 25	3 24	3 23	95	
.975	4 57	4 47	4 42	4 36	4 31	4 28	4 25	4 21	4 20	4 18	4 16	4 14	975	
.99	6 31	6 16	6 07	5 99	5 91	5 86	5 82	5 75	5 74	5 70	5 67	5 65	99	
.995	7.97	7 75	7 65	7 53	7 42	7 35	7 31	7 22	7 19	7 15	7 10	7 08	.995	
999	13 3	12 9	12 7	12 5	12 3	12 2	12 1	11 9	11 9	11 8	11 7	11 7	999	
9995	16 5	16 0	15 7	15 5	15 2	15 1	15 0	14 7	14 7	14 6	14 5	14 4	9995	
.0005	136	164	181	198	218	230	239	257	.262	271	281	287	0005	8
.001	155	184	200	218	238	250	259	277	282	292	300	306	001	
.005	214	244	261	279	299	311	319	337	341	351	358	364	005	
.01	250	281	297	315	334	346	354	372	376	385	392	398	01	
025	313	343	360	377	395	407	415	431	435	442	450	456	025	
.05	379	409	425	441	459	469	477	493	496	505	.510	516	05	
.10	472	500	515	531	547	556	563	578	581	588	595	599	10	
.25	684	707	718	730	743	751	756	767	769	775	780	783	25	
50	1 04	1 05	1 06	1 07	1 07	1.07	1 08	1 08	1 08	1 09	1.09	1.09	50	
.75	1 62	1 61	1 60	1 60	1 59	1 59	1 59	1 58	1 58	1 58	1 58	1.58	75	
.90	2 46	2 42	2 40	2 38	2 36	2 35	2 34	2 32	2 32	2 31	2 30	2.29	90	
.95	3 22	3 15	3 12	3 08	3 04	3 02	3 01	2 97	2 97	2 95	2.94	2.93	95	
.975	4 10	4 00	3 95	3 89	3 84	3 81	3 78	3 74	3 73	3 70	3.68	3.67	975	
.99	5 52	5 36	5 28	5 20	5 12	5 07	5 03	4 96	4 95	4 91	4 88	4.86	.99	
.995	6 81	6 61	6 50	6 40	6 29	6 22	6 18	6 09	6 06	6 02	5 98	5.95	.995	
.999	10 8	10 5	10 3	10 1	9 92	9 80	9 73	9 57	9 54	9 46	9 39	9.34	999	
.9995	13 1	12 7	12 5	12 2	12 0	11 8	11 8	11 6	11 5	11 4	11 4	11 3	9995	
.0005	141	171	188	207	228	242	251	270	276	287	297	.303	0005	9
.001	.160	.191	208	228	249	262	271	.291	296	307	316	323	001	
.005	.220	253	271	290	310	324	332	351	356	366	376	382	005	
.01	.257	289	307	.326	346	358	368	386	391	400	410	.415	01	
.025	320	352	370	388	408	420	428	446	450	.459	467	473	025	
05	.386	418	435	452	471	.483	490	508	510	518	526	532	05	
10	479	509	525	541	558	568	575	588	594	602	610	613	10	
25	.687	711	723	736	749	757	762	773	776	782	.787	791	25	
.50	1 03	1 04	1 05	1 05	1 06	1 06	1 07	1 07	1 07	1 08	1 08	1.08	.50	
.75	1 57	1 56	1 56	1 55	1 55	1 54	1 54	1 53	1 53	1 53	1 53	1.53	75	
.90	2 34	2 30	2 28	2 25	2 23	2 22	2 21	2 19	2 18	2 17	2 17	2.16	90	
.95	3 01	2 94	2 90	2 86	2 83	2 80	2 79	2 76	2 75	2 73	2 72	2 71	95	
975	3 77	3 67	3 61	3 56	3 51	3 47	3 45	3 40	3 39	3 37	3 35	3.33	975	
.99	4 96	4 81	4.73	4 65	4 57	4 52	4 48	4 42	4 40	4 36	4 33	4.31	99	
.995	6 03	5 83	5 73	5 62	5 52	5.45	5 41	5 32	5 30	5 26	5 21	5.19	.995	
.999	9.24	8 90	8 72	8 55	8 37	8 26	8 19	8 04	8 00	7 93	7.86	7.81	.999	
.9995	11 0	10 6	10.4	10 2	9 94	9.80	9 71	9 53	9 49	9 40	9.32	9.26	.9995	

Table A-9. Percentiles of the *F* distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

ν_2, DEGREES OF FREEDOM FOR DENOMINATOR	*Cum. prop.*	1	2	3	4	5	6	7	8	9	10	11	12	*Cum. prop.*
10	0005	$0^{6}41$	$0^{3}50$	$0^{2}49$	015	.028	043	057	071	085	097	108	119	0005
	001	$0^{5}17$	$0^{2}10$	$0^{2}77$	021	037	054	071	087	101	114	126	137	001
	005	$0^{4}41$	$0^{2}50$	023	048	073	098	119	139	156	171	185	197	005
	01	$0^{3}17$	010	037	069	100	127	151	172	.190	206	220	233	01
	.025	$0^{2}10$	025	069	113	151	183	210	.233	252	269	283	296	025
	.05	$0^{2}41$	052	114	168	211	246	275	.299	.319	336	351	363	05
	.10	.017	106	191	.255	303	340	370	.394	414	.430	444	457	10
	25	107	296	409	480	529	565	592	613	631	645	657	667	25
	50	.490	743	845	899	932	954	971	983	992	1 00	1 01	1 01	50
	75	1.49	1 60	1 60	1 59	1 59	1 58	1 57	1 56	1 56	1 55	1 55	1 54	75
	90	3 28	2 92	2 73	2 61	2 52	2 46	2 41	2 38	2 35	2 32	2 30	2 28	90
	.95	4 96	4 10	3 71	3 48	3 33	3 22	3 14	3 07	3 02	2 98	2 94	2.91	95
	.975	6 94	5 46	4 83	4 47	4 24	4 07	3 95	3 85	3 78	3.72	3 66	3.62	975
	.99	10 0	7 56	6 55	5 99	5 64	5 39	5 20	5 06	4 94	4 85	4 77	4.71	99
	.995	12 8	9 43	8 08	7 34	6 87	6 54	6 30	6 12	5 97	5 85	5 75	5.66	995
	.999	21 0	14 9	12 6	11 3	10 5	9 92	9 52	9 20	8 96	8 75	8 58	8 44	999
	9995	25 5	17 9	15 0	13.4	12 4	11 8	11 3	10 9	10 6	10 3	10.1	9.93	9995
11	0005	$0^{6}41$	$0^{3}50$	$0^{2}49$	015	028	043	058	072	086	099	111	121	0005
	.001	$0^{5}16$	$0^{2}10$	$0^{2}78$	021	038	055	072	088	103	116	129	140	001
	.005	$0^{4}40$	$0^{2}50$	023	048	074	099	121	141	158	174	188	200	005
	.01	$0^{3}16$	010	037	069	100	128	153	175	193	210	224	237	01
	025	$0^{2}10$	025	069	114	152	185	212	236	256	273	288	301	025
	05	$0^{2}41$	052	114	168	212	248	278	302	323	340	355	368	05
	10	017	106	192	256	305	342	373	397	417	435	448	461	10
	25	107	295	408	481	529	565	592	614	.633	645	658	667	25
	50	486	739	840	893	926	948	964	977	986	994	1 00	1 01	50
	75	1 47	1 58	1 58	1 57	1 56	1 55	1.54	1 53	1 53	1 52	1 52	1 51	75
	90	3 23	2 86	2 66	2 54	2 45	2 39	2 34	2 30	2 27	2 25	2 23	2 21	90
	95	4.84	3 98	3 59	3 36	3 20	3 09	3 01	2 95	2 90	2 85	2 82	2 79	95
	975	6.72	5 26	4 63	4 28	4 04	3 88	3 76	3 66	3 59	3 53	3 47	3.43	975
	99	9 65	7 21	6 22	5 67	5 32	5 07	4 89	4 74	4 63	4.54	4 46	4 40	99
	995	12 2	8 91	7 60	6 88	6 42	6 10	5 86	5 68	5 54	5 42	5 32	5 24	995
	999	19 7	13 8	11 6	10 3	9 58	9 05	8 66	8 35	8 12	7 92	7 76	7 62	999
	9995	23 6	16 4	13 6	12 2	11.2	10 6	10 1	9 76	9 48	9.24	9 04	8 88	9995
12	.0005	$0^{6}41$	$0^{3}50$	$0^{2}49$	015	028	.044	058	073	087	101	113	124	0005
	.001	$0^{5}16$	$.0^{2}10$	$0^{2}78$	021	038	056	073	089	104	118	131	143	001
	005	$0^{4}39$	$.0^{2}50$	023	048	075	100	.122	143	161	177	.191	.204	005
	01	$.0^{3}16$	.010	037	070	101	130	155	176	196	212	227	.241	01
	.025	$.0^{2}10$	025	.070	114	153	186	214	238	259	276	.292	305	025
	.05	$.0^{2}41$	052	114	169	214	250	280	305	325	343	358	372	05
	.10	016	106	192	257	306	344	.375	400	420	438	452	466	10
	.25	.106	295	408	480	530	566	594	.616	633	649	662	671	25
	.50	.484	735	835	888	921	943	.959	972	981	989	995	1 00	50
	.75	1.46	1 56	1 56	1.55	1 54	1 53	1 52	1 51	1 51	1 50	1 50	1 49	75
	.90	3 18	2 81	2 61	2 48	2 39	2 33	2 28	2 24	2 21	2 19	2 17	2 15	90
	95	4 75	3 89	3 49	3 26	3 11	3 00	2.91	2 85	2 80	2 75	2 72	2 69	95
	975	6 55	5 10	4 47	4.12	3 89	3.73	3 61	3 51	3 44	3 37	3 32	3 28	975
	.99	9 33	6 93	5 95	5.41	5 06	4 82	4 64	4 50	4 39	4 30	4 22	4 16	99
	.995	11 8	8 51	7 23	6.52	6 07	5 76	5 52	5.35	5 20	5 09	4 99	4 91	995
	999	18 6	13 0	10 8	9.63	8 89	8 38	8 00	7 71	7 48	7 29	7 14	7 01	999
	9995	22.2	15 3	12 7	11 2	10.4	9 74	9 28	8 94	8 66	8.43	8.24	8 08	9995

Table A-9. Percentiles of the F distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

Cum. prop.	15	20	24	30	40	50	60	100	120	200	500	∞	*Cum prop*	ν_2, DEGREES OF FREEDOM FOR DENOMINATOR
0005	145	177	195	215	238	251	262	282	288	299	311	319	0005	**10**
001	164	197	216	236	258	272	282	303	309	321	331	338	001	
005	226	260	279	299	321	334	344	365	370	.380	391	397	005	
01	263	297	316	336	357	370	380	400	405	415	424	431	01	
025	327	360	379	398	419	431	441	459	464	474	483	488	025	
05	393	426	444	462	481	493	502	518	523	532	541	546	05	
10	486	516	532	549	567	578	586	602	605	614	621	625	10	
25	691	714	727	740	754	762	767	779	782	788	793	797	25	
50	1 02	1 03	1 04	1 05	1 05	1 06	1 06	1 06	1 06	1 07	1 07	1 07	50	
75	1 53	1 52	1 52	1 51	1 51	1 50	1 50	1 49	1 49	1 49	1 48	1 48	75	
90	2 24	2 20	2 18	2 16	2 13	2 12	2 11	2 09	2 08	2 07	2 06	2 06	90	
95	2 85	2 77	2 74	2 70	2 66	2 64	2 62	2 59	2 58	2 56	2 55	2 54	95	
975	3 52	3 42	3 37	3 31	3 26	3 22	3 20	3 15	3 14	3 12	3 09	3 08	975	
99	4 56	4 41	4 33	4 25	4 17	4 12	4 08	4 01	4 00	3 96	3 93	3 91	99	
995	5 47	5 27	5 17	5 07	4 97	4 90	4 86	4 77	4 75	4 71	4.67	4 64	995	
999	8 13	7 80	7 64	7 47	7 30	7 19	7 12	6 98	6 94	6 87	6 81	6 76	999	
9995	9 56	9 16	8 96	8 75	8 54	8 42	8 33	8 16	8 12	8 04	7.96	7 90	9995	
0005	.148	182	201	222	246	261	271	293	299	312	324	331	0005	**11**
001	168	202	222	243	266	282	292	313	320	332	343	353	001	
005	231	266	286	308	330	345	355	376	382	394	403	412	005	
01	268	304	324	344	366	380	391	412	417	427	439	444	01	
025	332	368	386	.407	429	442	450	472	476	485	495	503	025	
05	398	433	452	469	490	503	513	529	535	543	552	559	05	
.10	490	524	541	559	578	588	595	614	617	625	633	637	10	
.25	.694	719	730	744	758	767	773	780	788	794	799	803	25	
.50	1 02	1 03	1 03	1 04	1 05	1 05	1 05	1 06	1 06	1 06	1 06	1 06	50	
75	1 50	1 49	1 49	1 48	1 47	1 47	1 47	1 46	1 46	1 46	1 45	1 45	75	
.90	2 17	2 12	2 10	2 08	2 05	2 04	2 03	2 00	2 00	1 99	1 98	1 97	90	
.95	2 72	2 65	2 61	2 57	2 53	2 51	2 49	2 46	2 45	2 43	2 42	2.40	95	
.975	3.33	3 23	3 17	3 12	3 06	3 03	3 00	2 96	2 94	2 92	2 90	2.88	975	
99	4 25	4 10	4 02	3 94	3 86	3 81	3 78	3 71	3 69	3 66	3 62	3.60	99	
.995	5 05	4 86	4 76	4 65	4 55	4 49	4 45	4 36	4 34	4 29	4 25	4.23	995	
.999	7 32	7 01	6 85	6 68	6 52	6 41	6 35	6 21	6 17	6 10	6 04	6.00	.999	
.9995	8 52	8 14	7 94	7 75	7 55	7 43	7 35	7.18	7 14	7 06	6.98	6.93	.9995	
.0005	152	.186	206	228	.253	269	280	305	311	.323	337	345	0005	**12**
.001	172	207	.228	250	275	291	302	.326	.332	344	357	.365	001	
005	235	.272	292	315	.339	355	365	388	393	405	417	424	.005	
.01	.273	310	330	352	.375	391	401	422	.428	441	450	.458	01	
.025	.337	.374	394	416	.437	450	461	.481	487	498	508	514	.025	
.05	.404	439	458	.478	.499	513	522	.541	545	.556	565	571	05	
.10	496	528	546	564	583	595	604	.621	.625	633	641	647	10	
.25	.695	721	734	748	.762	771	777	.789	792	.799	804	808	25	
.50	1 01	1 02	1 03	1 03	1 04	1 04	1 05	1 05	1 05	1 05	1 06	1 06	50	
.75	1 48	1 47	1 46	1 45	1 45	1 44	1 44	1.43	1.43	1.43	1 42	1 42	75	
.90	2 11	2 06	2 04	2 01	1 99	1 97	1 96	1 94	1 93	1 92	1 91	1 90	90	
.95	2 62	2 54	2 51	2 47	2 43	2 40	2 38	2 35	2 34	2.32	2 31	2 30	95	
.975	3 18	3 07	3 02	2 96	2 91	2 87	2.85	2 80	2 79	2.76	2.74	2.72	975	
.99	4 01	3 86	3 78	3 70	3 62	3 57	3 54	3 47	3 45	3.41	3.38	3.36	99	
.995	4 72	4 53	4 43	4 33	4 23	4.17	4 12	4 04	4 01	3.97	3.93	3.90	995	
.999	6 71	6 40	6 25	6 09	5 93	5 83	5.76	5 63	5 59	5.52	5.46	5.42	999	
.9995	7.74	7.37	7.18	7 00	6 80	6 68	6.61	6 45	6 41	6 33	6.25	6.20	.9995	

Table A-9. Percentiles of the F distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

ν_2, DEGREES OF FREEDOM FOR DENOMINATOR	*Cum. prop.*	1	2	3	4	5	6	7	8	9	10	11	12	*Cum. prop*
15	.0005	$0^{6}41$	$0^{3}50$	$.0^{2}49$	015	029	045	061	076	091	105	.117	129	0005
	001	$.0^{5}16$	$0^{2}10$	$0^{2}79$	021	039	057	075	092	108	123	137	149	001
	.005	$.0^{4}39$	$0^{2}50$	023	049	076	102	125	147	166	183	198	212	005
	.01	$0^{3}16$	010	037	070	103	132	158	181	202	219	235	249	01
	.025	$.0^{2}10$	025	070	116	156	190	219	.244	265	284	300	315	025
	.05	$0^{2}41$	051	115	170	216	254	.285	311	333	.351	368	382	05
	.10	.016	106	192	258	309	.348	380	.406	427	446	461	475	10
	25	105	293	407	480	531	.568	596	618	637	.652	667	676	25
	50	478	726	826	878	911	933	948	960	970	.977	.984	989	50
	75	1 43	1 52	1.52	1 51	1.49	1 48	1 47	1 46	1 46	1 45	1 44	1 44	75
	.90	3 07	2 70	2.49	2 36	2 27	2 21	2.16	2 12	2 09	2 06	2 04	2 02	90
	.95	4 54	3.68	3 29	3 06	2 90	2 79	2 71	2 64	2 59	2.54	2.51	2 48	.95
	975	6 20	4.76	4.15	3 80	3 58	3 41	3 29	3 20	3.12	3 06	3 01	2 96	975
	99	8.68	6 36	5.42	4.89	4 56	4 32	4 14	4.00	3.89	3 80	3 73	3 67	99
	.995	10 8	7.70	6 48	5 80	5 37	5 07	4 85	4.67	4.54	4.42	4.33	4 25	995
	.999	16.6	11.3	9 34	8 25	7 57	7 09	6 74	6 ·47	6 26	6 08	5 93	5 81	999
	9995	19.5	13.2	10.8	9 48	8 66	8 10	7 68	7 36	7.11	6 91	6 75	6 60	9995
20	.0005	$.0^{6}40$	$.0^{3}50$	$0^{2}50$	015	029	.046	.063	079	094	109	.123	.136	0005
	.001	$.0^{5}16$	$.0^{2}10$	$.0^{2}79$	.022	.039	.058	.077	.095	112	128	.143	.156	001
	.005	$.0^{4}39$	$.0^{2}50$	023	050	077	.104	.129	151	.171	190	.206	.221	.005
	.01	$.0^{3}16$	010	.037	.071	.105	.135	.162	187	208	227	.244	.259	01
	.025	$.0^{2}10$	025	.071	.117	158	193	.224	250	273	292	310	.325	.025
	.05	$.0^{2}40$	.051	.115	.172	219	.258	290	.318	340	360	377	.393	05
	.10	.016	.106	.193	.260	312	353	385	412	435	454	472	.485	10
	.25	.104	.292	.407	480	531	569	598	622	.641	656	.671	681	25
	.50	.472	.718	816	.868	900	922	938	950	959	966	972	.977	50
	.75	1 40	1.49	1.48	1.47	1.45	1.44	1 43	1 42	1.41	1.40	1.39	1 39	75
	.90	2 97	2.59	2 38	2.25	2.16	2.09	2.04	2.00	1 96	1 94	1.91	1 89	90
	.95	4.35	3.49	3.10	2.87	2.71	2 60	2 51	2 45	2.39	2.35	2.31	2 28	95
	.975	5 87	4 46	3.86	3.51	3.29	3.13	3 01	2 91	2 84	2 77	2.72	2 68	975
	.99	8 10	5.85	4.94	4 43	4 10	3 87	3 70	3.56	3 46	3.37	3 29	3 23	99
	.995	9 94	6 99	5 82	5 17	4.76	4.47	4 26	4.09	3 96	3.85	3.76	3 68	995
	.999	14 8	9 95	8.10	7 10	6 46	6 02	5 69	5 44	5 24	5 08	4.94	4 82	999
	9995	17.2	11.4	9 20	8 02	7 28	6 76	6 38	6 08	5 85	5 66	5 51	5 38	9995
24	.0005	$.0^{6}40$	$.0^{3}50$	$.0^{2}50$	015	030	046	.064	080	096	112	126	139	.0005
	.001	$.0^{5}16$	$.0^{2}10$	$0^{2}79$	.022	040	059	079	097	115	131	146	160	001
	.005	$.0^{4}40$	$0^{2}50$	.023	050	078	106	131	.154	.175	193	210	226	.005
	.01	$.0^{3}16$	010	.038	.072	106	137	165	.189	211	231	.249	264	01
	.025	$.0^{2}10$	.025	.071	117	.159	195	227	253	.277	.297	315	331	025
	.05	$.0^{2}40$	.051	.116	.173	221	260	293	.321	345	.365	.383	399	.05
	.10	.016	.106	.193	261	.313	355	388	416	.439	.459	476	491	10
	.25	.104	.291	.406	.480	.532	570	600	.623	.643	659	671	.684	25
	.50	.469	.714	.812	.863	895	.917	932	944	.953	961	967	972	.50
	.75	1.39	1 47	1.46	1.44	1.43	1 41	1 40	1 39	1 38	1.38	1 37	1 36	75
	.90	2.93	2.54	2.33	2 19	2.10	2 04	1.98	1 94	1 91	1.88	1 85	1 83	.90
	.95	4.26	3.40	3.01	2 78	2.62	2.51	2 42	2.36	2 30	2 25	2.21	2.18	.95
	.975	5.72	4 32	3.72	3 38	3.15	2 99	2 87	2 78	2 70	2 64	2 59	2.54	975
	99	7.82	5.61	4.72	4.22	3.90	3.67	3 50	3 36	3.26	3 17	3.09	3.03	.99
	.995	9.55	6 66	5.52	4 89	4.49	4.20	3 99	3 83	3 69	3 59	3.50	3 42	.995
	.999	14 0	9.34	7.55	6 59	5.98	5.55	5 23	4 99	4 80	4 64	4 50	4.39	999
	.9995	16.2	10.6	8.52	7 39	6.68	6 18	5.82	5 54	5.31	5 13	4.98	4.85	9995

Table A-9. Percentiles of the *F* distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

Cum. prop.	15	20	24	30	40	50	60	100	120	200	500	∞	*Cum prop.*	ν_2, DEGREES OF FREEDOM FOR DENOMINATOR
0005	159	197	220	244	272	290	303	330	339	353	368	377	0005	**15**
.001	181	219	242	266	294	313	325	352	360	375	388	398	001	
.005	246	286	308	333	360	377	389	415	422	435	448	457	005	
.01	284	324	346	370	397	413	425	450	456	469	483	490	01	
.025	.349	389	410	433	.458	474	485	.508	514	526	538	546	025	
.05	.416	454	474	496	519	535	545	565	571	581	592	600	05	
.10	507	542	561	581	602	614	.624	.641	647	658	667	672	.10	
.25	.701	728	742	757	.772	782	788	802	805	812	818	822	.25	
.50	1.00	1 01	1 02	1 02	1 03	1 03	1 03	1 04	1 04	1 04	1 04	1.05	50	
.75	1 43	1.41	1 41	1 40	1.39	1 39	1 38	1 38	1 37	1 37	1 36	1.36	75	
.90	1 97	1.92	1 90	1 87	1 85	1 83	1 82	1 79	1 79	1 77	1 76	1.76	90	
.95	2 40	2 33	2 39	2 25	2 20	2 18	2.16	2 12	2 11	2 10	2 08	2.07	95	
.975	2 86	2 76	2 70	2 64	2 59	2 55	2 52	2 47	2 46	2 44	2 41	2.40	975	
.99	3.52	3.37	3 29	3 21	3 13	3 08	3 05	2 98	2 96	2 92	2 89	2.87	99	
.995	4.07	3 88	3 79	3 69	3 59	3 52	3.48	3 39	3 37	3 33	3 29	3 26	995	
.999	5 54	5.25	5 10	4.95	4 80	4.70	4 64	4.51	4 47	4 41	4 35	4.31	.999	
.9995	6.27	5.93	5 75	5.58	5 40	5 29	5 21	5 06	5 02	4 94	4 87	4.83	9995	
.0005	169	211	.235	263	.295	316	331	364	375	391	408	422	.0005	**20**
.001	.191	.233	.258	286	.318	339	354	.386	395	413	429	441	.001	
.005	258	301	.327	354	.385	.405	419	.448	457	474	490	500	005	
.01	.297	340	.365	.392	.422	441	.455	483	.491	508	521	532	01	
.025	.363	406	.430	.456	.484	503	514	.541	548	562	575	585	025	
.05	.430	471	.493	518	.544	562	572	.595	.603	617	629	637	05	
.10	.520	557	.578	.600	.623	637	648	671	675	685	694	704	10	
.25	.708	.736	751	.767	.784	794	801	816	820	827	835	840	25	
.50	.989	i 00	1 01	1.01	1 02	1 02	1 02	1 03	1 03	1 03	1 03	1.03	.50	
.75	1.37	1 36	1 35	1 34	1 33	1 33	1.32	1 31	1 31	1 30	1 30	1.29	.75	
.90	1 84	1.79	1 77	1.74	1.71	1 69	1.68	1 65	1 64	1 63	1 62	1.61	90	
.95	2.20	2.12	2 08	2 04	1.99	1.97	1 95	1 91	1 90	1 88	1 86	1.84	.95	
.975	2.57	2.46	2.41	2 35	2 29	2 25	2.22	2.17	2 16	2 13	2.10	2.09	.975	
.99	3 09	2 94	2 86	2 78	2 69	2 64	2 61	2 54	2 52	2 48	2 44	2.42	99	
.995	3.50	3 32	3 22	3 12	3 02	2 96	2 92	2 83	2 81	2.76	2 72	2.69	995	
.999	4.56	4.29	4.15	4.01	3 86	3 77	3 70	3 58	3 54	3.48	3 42	3.38	.999	
.9995	5 07	4.75	4.58	4 42	4.24	4.15	4.07	3.93	3 90	3 82	3 75	3.70	.9995	
.0005	.174	.218	.244	274	.309	331	.349	384	395	416	434	449	0005	**24**
.001	.196	241	.268	.298	.332	354	371	405	.417	.437	455	469	.001	
.005	.264	310	.337	.367	.400	422	437	469	.479	498	515	.527	005	
.01	.304	.350	.376	405	437	459	.473	505	.513	529	546	558	.01	
.025	.370	415	.441	.468	.498	518	531	562	568	585	599	610	025	
.05	.437	480	504	.530	.558	575	588	613	622	637	649	.659	05	
.10	.527	566	.588	.611	.635	651	662	685	.691	704	715	723	.10	
.25	.712	.741	.757	773	.791	.802	809	825	.829	837	844	.850	25	
.50	.983	.994	1 00	1.01	1 01	1.02	1.02	1 02	1 02	1 02	1 03	1.03	.50	
.75	1.35	1.33	1.32	1.31	1 30	1.29	1 29	1 28	1 28	1.27	1.27	1.26	75	
.90	1.78	1.73	1.70	1.67	1 64	1.62	1 61	1.58	1 57	1.56	1 54	1.53	90	
.95	2.11	2.03	1.98	1 94	1 89	1.86	1 84	1 80	1 79	1 77	1 75	1.73	.95	
.975	2.44	2 33	2 27	2.21	2.15	2.11	2 08	2 02	2 01	1 98	1 95	1.94	975	
.99	2 89	2.74	2 66	2 58	2 49	2 44	2 40	2 33	2 31	2 27	2 24	2.21	99	
.995	3 25	3 06	2.97	2 87	2 77	2 70	2 66	2 57	2 55	2 50	2 46	2.43	.995	
.999	4 14	3 87	3 74	3.59	3 45	3 35	3 29	3 16	3 14	3 07	3 01	2.97	999	
.9995	4 55	4 25	4 09	3 93	3 76	3 66	3 59	3 44	3 41	3 33	3 27	3.22	.9995	

Table A-9. Percentiles of the *F* distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

ν_2, DEGREES OF FREEDOM FOR DENOMINATOR	*Cum. prop*	1	2	3	4	5	6	7	8	9	10	11	12	*Cum. prop.*
30	0005	$0^{6}40$	$0^{3}50$	$0^{2}50$	015	030	047	065	082	098	114	129	143	0005
	001	$0^{5}16$	$0^{2}10$	$0^{2}80$	022	040	060	080	099	117	134	150	164	001
	005	$0^{4}40$	$0^{2}50$	024	050	079	107	133	156	178	.197	215	231	005
	01	$0^{3}16$	010	038	072	107	138	167	192	215	235	254	270	01
	025	$0^{2}10$	025	.071	118	161	197	229	257	281	302	321	337	.025
	.05	$0^{2}40$	051	116	174	222	263	296	325	349	370	389	.406	05
	10	016	106	193	262	315	357	391	420	443	.464	481	497	10
	25	103	290	406	480	532	571	601	625	645	661	676	688	25
	50	466	709	807	858	.890	912	927	939	948	955	961	966	50
	.75	1 38	1 45	1 44	1 42	1 41	1 39	1 38	1 37	1 36	1 35	1 35	1.34	75
	90	2 88	2 49	2 28	2 14	2 05	1 98	1 93	1 88	1.85	1 82	1 79	1.77	90
	.95	4.17	3 32	2 92	2 69	2 53	2 42	2 33	2 27	2 21	2 16	2 13	2.09	95
	.975	5 57	4 18	3 59	3 25	3 03	2.87	2 75	2 65	2 57	2 51	2 46	2 41	975
	.99	7.56	5 39	4 51	4 02	3 70	3.47	3 30	3 17	3 07	2 98	2 91	2.84	99
	.995	9.18	6 35	5 24	4 62	4 23	3 95	3 74	3 58	3 45	3 34	3 25	3 18	995
	.999	13 3	8 77	7 05	6 12	5 53	5 12	4 82	4 58	4 39	4.24	4 11	4 00	999
	9995	15 2	9 90	7 90	6.82	6.14	5 66	5 31	5 04	4 82	4 65	4 51	4 38	9995
40	.0005	$0^{6}40$	$0^{3}50$	$0^{2}50$	016	030	048	066	084	100	.117	132	.147	0005
	.001	$.0^{5}16$	$0^{2}10$	$0^{2}80$	022	042	.061	081	101	119	137	.153	.169	001
	.005	$.0^{4}40$	$0^{2}50$	024	.051	080	108	135	159	181	201	.220	.237	005
	01	$0^{3}16$	010	038	.073	108	140	169	195	219	240	259	276	01
	.025	$.0^{3}99$	025	071	119	162	.199	232	260	.285	307	.327	.344	025
	.05	$0^{2}40$	051	116	175	224	265	299	329	.354	376	.395	.412	05
	.10	016	106	194	.263	317	360	394	424	448	.469	.488	.504	10
	25	103	290	.405	480	533	572	603	627	647	.664	680	.691	25
	50	463	705	.802	854	885	907	922	934	943	.950	.956	.961	50
	.75	1.36	1 44	1 42	1.40	1.39	1 37	1.36	1.35	1 34	1 33	1 32	1 31	75
	.90	2 84	2 44	2 23	2 09	2 00	1 93	1 87	1 83	1 79	1 76	1 73	1.71	90
	.95	4.08	3 23	2.84	2 61	2 45	2 34	2 25	2.18	2 12	2.08	2 04	2 00	95
	.975	5 42	4 05	3 46	3 13	2 90	2 74	2 62	2 53	2 45	2 39	2 33	2 29	975
	.99	7 31	5 18	4 31	3 83	3 51	3 29	3 12	2 99	2.89	2.80	2 73	2.66	99
	.995	8 83	6 07	4 98	4 37	3 99	3.71	3.51	3 35	3 22	3.12	3 03	2 95	995
	.999	12 6	8 25	6 60	5 70	5 13	4 73	4.44	4 21	4 02	3 87	3 75	3 64	999
	.9995	14 4	9 25	7 33	6 30	5 64	5 19	4.85	4 59	4 38	4 21	4 07	3 95	9995
60	.0005	$.0^{6}40$	$0^{3}50$	$.0^{2}51$	016	.031	.048	067	085	103	.120	.136	152	.0005
	.001	$0^{5}16$	$0^{2}10$	$0^{2}80$	022	041	062	083	103	122	.140	.157	174	.001
	.005	$0^{4}40$	$0^{2}50$	024	051	081	110	137	.162	185	206	225	243	005
	.01	$0^{3}16$	010	038	073	109	142	172	199	223	245	265	283	01
	.025	$.0^{3}99$	.025	071	120	163	202	235	264	290	313	.333	351	.025
	.05	$.0^{2}40$	051	116	176	226	267	303	333	359	.382	402	419	05
	.10	016	.106	.194	264	318	362	398	428	453	475	493	510	10
	25	102	.289	405	480	534	.573	604	629	.650	667	.680	695	.25
	.50	461	701	798	.849	880	901	917	928	937	945	951	956	50
	75	1 35	1 42	1.41	1 38	1 37	1 35	1 33	1 32	1 31	1 30	1.29	1 29	.75
	90	2.79	2 39	2.18	2 04	1 95	1.87	1 82	1 77	1 74	1.71	1.68	1 66	.90
	.95	4 00	3 15	2 76	2 53	2 37	2 25	2 17	2 10	2 04	1.99	1.95	1 92	.95
	.975	5 29	3 93	3 34	3 01	2 79	2 63	2 51	2.41	2.33	2.27	2.22	2 17	.975
	.99	7 08	4 98	4 13	3 65	3 34	3 12	2 95	2.82	2 72	2.63	2.56	2.50	99
	.995	8 49	5 80	4 73	4 14	3 76	3 49	3 29	3 13	3 01	2 90	2.82	2 74	.995
	999	12.0	7 76	6 17	5 31	4 76	4 37	4 09	3 87	3 69	3 54	3.43	3 31	999
	.9995	13.6	8 65	6 81	5 82	5 20	4 76	4 44	4 18	3.98	3 82	3.69	3 57	9995

Table A-9. Percentiles of the F distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

Cum. prop.	15	20	24	30	40	50	60	100	120	200	500	∞	*Cum prop*	ν_2, DEGREES OF FREEDOM FOR DENOMINATOR
0005	.179	226	254	.287	325	350	,369	.410	420	.444	467	483	0005	**30**
001	.202	250	278	311	348	373	391	431	442	465	488	503	001	
.005	.271	320	.349	381	.416	441	457	495	504	524	543	559	005	
.01	.311	.360	388	419	.454	476	493	529	538	.559	575	590	01	
.025	.378	.426	453	.482	515	535	551	585	592	610	625	639	025	
.05	.445	.490	.516	.543	.573	592	606	637	.644	.658	676	685	05	
.10	.534	575	.598	.623	.649	667	678	.704	710	725	.735	746	10	
.25	.716	746	763	780	798	810	818	.835	839	848	856	862	25	
.50	.978	989	994	1 00	1 01	1 01	1 01	1 02	1 02	1.02	1 02	1 02	50	
.75	1 32	1 30	1 29	1 28	1 27	1 26	1 26	1 25	1 24	1.24	1 23	1 23	75	
.90	1 72	1 67	1 64	1.61	1 57	1.55	1 54	1.51	1 50	1.48	1 47	1 46	90	
.95	2 01	1 93	1 89	1 84	1 79	1.76	1 74	1 70	1 68	1 66	1 64	1 62	95	
.975	2 31	2 20	2 14	2 07	2 01	1 97	1 94	1 88	1 87	1 84	1 81	1 79	975	
.99	2 70	2 55	2 47	2 39	2 30	2 25	2 21	2 13	2 11	2 07	2 03	2 01	99	
.995	3 01	2 82	2 73	2.63	2 52	2.46	2.42	2 32	2 30	2 25	2 21	2.18	995	
.999	3.75	3 49	3 36	3 22	3 07	2.98	2 92	2 79	2 76	2 69	2 63	2.59	999	
.9995	4 10	3 80	3 65	3 48	3.32	3 22	3.15	3.00	2 97	2.89	2 82	2 78	9995	
.0005	.185	236	266	301	343	373	393	.441	453	480	504	525	0005	**40**
.001	.209	.259	290	326	.367	396	415	461	473	.500	524	545	001	
.005	.279	.331	362	.396	.436	.463	481	.524	.534	.559	581	599	005	
.01	.319	371	.401	435	.473	.498	.516	556	567	.592	613	628	01	
.025	.387	437	466	498	.533	.556	573	.610	620	.641	662	.674	025	
.05	.454	502	529	.558	.591	613	627	.658	669	.685	704	717	05	
.10	542	585	609	.636	.664	683	696	.724	731	.747	762	772	10	
.25	720	752	769	787	806	819	828	846	.851	.861	870	877	25	
.50	972	983	989	994	1.00	1.00	1.01	1 01	1 01	1.01	1 02	1 02	50	
.75	1 30	1 28	1 26	1 25	1.24	1.23	1.22	1 21	1 21	1 20	1.19	1 19	75	
.90	1.66	1.61	1 57	1 54	1.51	1 48	1.47	1 43	1 42	1.41	1 39	1 38	90	
.95	1.92	1 84	1 79	1 74	1 69	1.66	1 64	1.59	1.58	1 55	1 53	1.51	95	
.975	2 18	2 07	2 01	1 94	1 88	1.83	1.80	1 74	1 72	1 69	1 66	1 64	975	
.99	2.52	2.37	2 29	2 20	2 11	2 06	2.02	1 94	1 92	1.87	1 83	1 80	99	
.995	2 78	2.60	2 50	2 40	2 30	2 23	2 18	2 09	2 06	2 01	1 96	1 93	.995	
.999	3.40	3.15	3 01	2 87	2.73	2 64	2 57	2.44	2 41	2 34	2 28	2 23	999	
.9995	3 68	3 39	3.24	3 08	2 92	2 82	2.74	2 60	2 57	2 49	2.41	2 37	9995	
.0005	192	246	278	318	.365	398	421	.478	493	527	561	585	0005	**60**
.001	216	270	304	343	.389	.421	444	.497	.512	545	579	602	.001	
.005	287	343	376	414	458	488	510	.559	572	602	633	652	005	
.01	.328	383	.416	.453	495	524	545	.592	604	633	.658	679	01	
.025	.396	450	481	515	555	581	600	.641	654	.680	704	720	025	
.05	.463	514	.543	.575	611	633	652	.690	700	.719	746	759	.05	
.10	.550	596	622	.650	682	.703	717	.750	758	.776	793	.806	10	
.25	.725	.758	776	.796	.816	830	840	860	865	877	888	896	25	
.50	.967	978	983	.989	.994	998	1.00	1 00	1 01	1 01	1 01	1.01	50	
.75	1 27	1 25	1 24	1 22	1 21	1.20	1 19	1 17	1 17	1.16	1 15	1.15	75	
.90	1.60	1 54	1.51	1 48	1 44	1.41	1 40	1 36	1 35	1.33	1 31	1.29	90	
.95	1.84	1 75	1 70	1 65	1 59	1.56	1 53	1 48	1 47	1 44	1 41	1.39	95	
.975	2 06	1 94	1 88	1 82	1.74	1.70	1 67	1.60	1.58	1 54	1 51	1.48	975	
.99	2 35	2 20	2 12	2 03	1 94	1.88	1 84	1 75	1.73	1 68	1.63	1.60	99	
.995	2 57	2 39	2 29	2 19	2 08	2 01	1 96	1 86	1 83	1 78	1.73	1.69	995	
.999	3 08	2 83	2.69	2 56	2 41	2.31	2 25	2 11	2 09	2 01	1 93	1.89	999	
.9995	3.30	3 02	2 87	2.71	2 55	2 45	2 38	2 23	2.19	2 11	2 03	1.98	9995	

Table A-9. Percentiles of the F distribution (*Continued*).

ν_1, DEGREES OF FREEDOM FOR NUMERATOR

ν_2, DEGREES OF FREEDOM FOR DENOMINATOR	*Cum. prop.*	1	2	3	4	5	6	7	8	9	10	11	12	*Cum. prop.*
120	0005	$0^{6}40$	$0^{3}50$	$0^{2}51$	016	031	049	067	087	105	123	140	156	0005
	001	$.0^{5}16$	$0^{2}10$	$0^{2}81$	023	042	063	.084	105	125	144	.162	179	001
	005	$.0^{4}39$	$0^{2}50$	024	051	081	111	139	165	189	211	230	249	005
	01	$0^{3}16$	010	038	074	110	143	.174	202	227	250	271	290	01
	.025	$.0^{3}99$	025	072	120	165	204	.238	.268	295	318	340	359	025
	05	$.0^{2}39$	.051	.117	177	227	270	.306	.337	364	388	408	.427	05
	.10	.016	105	194	265	320	365	.401	.432	458	.480	500	518	10
	.25	.102	288	.405	481	534	.574	.606	.631	652	670	685	699	25
	50	458	.697	.793	.844	875	896	912	923	932	939	945	.950	50
	75	1 34	1 40	1 39	1 37	1.35	1 33	1 31	1.30	1 29	1 28	1.27	1 26	75
	90	2 75	2 35	2 13	1.99	1 90	1 82	1 77	1.72	1 68	1 65	1.62	1 60	.90
	95	3 92	3 07	2.68	2 45	2 29	2 18	2 09	2 02	1 96	1 91	1.87	1 83	95
	975	5.15	3 80	3 23	2 89	2 67	2.52	2 39	2 30	2 22	2 16	2.10	2 05	975
	99	6 85	4 79	3.95	3 48	3 17	2.96	2 79	2 66	2 56	2.47	2.40	2 34	99
	995	8 18	5 54	4 50	3 92	3 55	3 28	3 09	2 93	2 81	2.71	2 62	2 54	995
	999	11 4	7 32	5 79	4 95	4 42	4 04	3 77	3 55	3.38	3.24	3 12	3 02	999
	9995	12 8	8 10	6 34	5 39	4 79	4 37	4 07	3 82	3.63	3 47	3.34	3 22	9995
∞	0005	$0^{6}39$	$0^{3}50$	$0^{2}51$	016	032	050	069	088	108	127	144	161	.0005
	001	$0^{5}16$	$0^{2}10$	$0^{2}81$	023	042	063	085	.107	.128	148	167	185	001
	005	$0^{4}39$	$0^{2}50$	024	052	082	113	141	168	193	216	236	256	005
	01	$0^{3}16$	010	038	074	111	145	177	.206	232	256	278	.298	01
	025	$0^{3}98$	025	072	.121	166	.206	241	.272	.300	325	347	367	.025
	05	$0^{2}39$	.051	117	.178	229	.273	310	342	.369	.394	.417	436	05
	10	016	105	195	266	.322	367	405	436	.463	487	508	525	10
	25	102	.288	.404	.481	.535	576	608	.634	.655	.674	690	703	.25
	50	455	693	789	839	870	891	907	.918	927	934	939	945	.50
	.75	1.32	1.39	1 37	1.35	1 33	1 31	1 29	1 28	1.27	1 25	1.24	1 24	.75
	.90	2.71	2 30	2.08	1.94	1.85	1 77	1 72	1 67	1 63	1.60	1.57	1 55	.90
	95	3 84	3 00	2.60	2.37	2 21	2 10	2 01	1 94	1 88	1.83	1.79	1 75	.95
	975	5.02	3 69	3.12	2.79	2.57	2.41	2 29	2 19	2 11	2 05	1.99	1 94	.975
	99	6.63	4 61	3.78	3 32	3.02	2 80	2 64	2 51	2 41	2 32	2 25	2 18	99
	995	7.88	5.30	4.28	3.72	3.35	3 09	2 90	2 74	2 62	2.52	2 43	2 36	995
	999	10 8	6.91	5 42	4 62	4.10	3 74	3 47	3 27	3 10	2 96	2.84	2.74	999
	9995	12 1	7.60	5.91	5 00	4.42	4 02	3 72	3 48	3 30	3 14	3 02	2 90	.9995

For sample sizes larger than, say, 30, a fairly good approximation to the F distribution percentiles can be obtained from

$$\log F_\alpha(\nu_1,\nu_2) \approx \left(\frac{a}{\sqrt{h - b}}\right) - cg$$

where $h = 2\nu_1\nu_2/(\nu_1 + \nu_2)$, $g = (\nu_2 - \nu_1)/\nu_1\nu_2$, and a, b, c are functions of α given below:

VALUES OF α

	α = .50	.75	.90	95	.975	99	995	.999	9995
a	0	.5859	1.1131	1.4287	1.7023	2.0206	2 2373	2.6841	2 8580
b	—	.58	.77	95	1.14	1.40	1.61	2.09	2 30
c	.290	.355	.527	.681	.846	1 073	1.250	1.672	1 857

Table A-9. Percentiles of the F distribution (*Continued*).

	ν_1, DEGREES OF FREEDOM FOR NUMERATOR													
Cum. prop.	15	20	24	30	40	50	60	100	120	200	500	∞	*Cum. prop.*	ν_2, DEGREES OF FREEDOM FOR DENOMINATOR
.0005	199	256	293	.338	390	429	458	524	543	578	614	676	0005	**120**
001	223	282	319	363	415	453	480	542	568	595	631	691	001	
005	.297	356	393	434	484	520	545	605	.623	661	702	733	005	
.01	338	397	433	474	522	556	.579	636	.652	688	725	755	01	
.025	.406	464	498	536	580	611	.633	684	.698	729	762	789	025	
.05	473	.527	559	594	634	661	682	727	.740	767	785	819	05	
.10	560	609	636	667	.702	726	.742	.781	.791	815	838	855	.10	
25	730	765	784	805	.828	843	.853	877	.884	897	911	923	25	
.50	961	972	978	983	989	992	.994	1 00	1 00	1 00	1.01	1.01	50	
75	1 24	1 22	1 21	1 19	1 18	1.17	1 16	1 14	1 13	1 12	1.11	1.10	75	
90	1 55	1 48	1 45	1 41	1 37	1.34	1 32	1 27	1 26	1 24	1.21	1.19	90	
.95	1 75	1.66	1 61	1 55	1 50	1.46	1 43	1.37	1 35	1 32	1 28	1.25	95	
975	1 95	1 82	1 76	1.69	1 61	1.56	1.53	1 45	1 43	1 39	1 34	1.31	975	
.99	2 19	2 03	1 95	1 86	1 76	1.70	1 66	1 56	1 53	1.48	1 42	1.38	99	
.995	2 37	2 19	2 09	1 98	1 87	1.80	1 75	1 64	1 61	1 54	1.48	1.43	995	
999	2 78	2 53	2 40	2 26	2 11	2.02	1 95	1 82	1 76	1 70	1 62	1.54	.999	
9995	2 96	2 67	2 53	2 38	2 21	2.11	2 01	1 88	1 84	1 75	1 67	1.60	9995	
0005	207	270	311	360	422	469	505	599	.624	704	804	1.00	0005	∞
.001	232	296	338	386	.448	493	.527	617	.649	719	819	1.00	001	
005	307	372	412	460	518	559	.592	671	699	762	.843	1 00	.005	
.01	349	413	452	499	554	595	625	699	.724	782	858	1 00	01	
025	.418	480	517	560	611	645	.675	741	.763	813	878	1 00	025	
05	.484	543	577	617	663	694	720	781	797	840	896	1 00	05	
.10	570	622	652	.687	.726	752	774	826	838	877	919	1.00	10	
25	736	.773	.793	.816	842	860	872	901	.910	.932	957	1 00	25	
50	.956	967	972	978	.983	.987	.989	993	994	.997	999	1 00	50	
.75	1.22	1.19	1.18	1 16	1.14	1.13	1 12	1 09	1.08	1 07	1.04	1 00	75	
.90	1.49	1.42	1 38	1.34	1.30	1.26	1 24	1 18	1.17	1 13	1.08	1 00	90	
95	1.67	1.57	1 52	1.46	1 39	1.35	1 32	1 24	1 22	1.17	1.11	1 00	95	
.975	1.83	1 71	1 64	1.57	1 48	1.43	1 39	1.30	1.27	1 21	1.13	1 00	.975	
99	2 04	1.88	1 79	1 70	1.59	1 52	1 47	1.36	1.32	1 25	1.15	1 00	99	
.995	2 19	2 00	1 90	1 79	1 67	1.59	1 53	1 40	1.36	1 28	1 17	1 00	995	
.999	2 51	2.27	2 13	1 99	1 84	1 73	1 66	1 49	1.45	1 34	1.21	1 00	.999	
9995	2.65	2.37	2 22	2 07	1 91	1.79	1 71	1.53	1.48	1 36	1 22	1.00	.9995	

The values given in this table are abstracted with permission from the following sources:

1. All values for ν_1, ν_2 equal to 50, 100, 200, 500 are from A. Hald, *Statistical Tables and Formulas*, John Wiley & Sons, Inc , New York, 1952.

2. For cumulative proportions .5, .75, .9, .95, .975, .99, 995 most of the values are from M Merrington and C. M. Thompson, *Biometrika*, vol 33 (1943), p. 73.

3. For cumulative proportions .999 the values are from C Colcord and L. S. Deming, *Sankhyā*, vol 2 (1936), p. 423.

4. For cum. prop. $= \alpha < .5$ the values are the reciprocals of values for $1 - \alpha$ (with ν_1 and ν_2 interchanged). The values in Merrington and Thompson and in Colcord and Deming are to five significant figures, and it is hoped (but not expected) that the reciprocals are correct as given. The values in Hald are to three significant figures, and the reciprocals are probably accurate within one to two digits in the third significant figure except for those values very close to unity, where they may be off four to five digits in the third significant figure

5. Gaps remaining in the table after using the above sources were filled in by interpolation.

$$\alpha = \frac{(\nu_1/\nu_2)^{\frac{1}{2}\nu_1}}{\beta(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)} \int_{-\infty}^{F\alpha} F^{\frac{1}{2}\nu_1 - 1} \left(1 + \frac{\nu_1 F}{\nu_2}\right)^{-(\nu_1+\nu_2)/2} dF$$

Table A-10. Critical values for Cochran's test of equal variance.

Values given are for the statistic (largest s^2)/(Σs_i^2), where each of the k values of s^2 has ν degrees of freedom.

PERCENTILE 95

k \ ν	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8332	0.8159	0.8010	0.7880	0.7341	0.6602	0.5813	0.5000
3	0.9669	0.8709	0.7977	0.7457	0.7071	0.6771	0.6530	0.6333	0.6167	0.6025	0.5466	0.4748	0.4031	0.3333
4	0.9065	0.7679	0.6841	0.6287	0.5895	0.5598	0.5365	0.5175	0.5017	0.4884	0.4366	0.3720	0.3093	0.2500
5	0.8412	0.6838	0.5981	0.5441	0.5065	0.4783	0.4564	0.4387	0.4241	0.4118	0.3645	0.3066	0.2513	0.2000
6	0.7808	0.6161	0.5321	0.4803	0.4447	0.4184	0.3980	0.3817	0.3682	0.3568	0.3135	0.2612	0.2119	0.1667
7	0.7271	0.5612	0.4800	0.4307	0.3974	0.3726	0.3535	0.3384	0.3259	0.3154	0.2756	0.2278	0.1833	0.1429
8	0.6798	0.5157	0.4377	0.3910	0.3595	0.3362	0.3185	0.3043	0.2926	0.2829	0.2462	0.2022	0.1616	0.1250
9	0.6385	0.4775	0.4027	0.3584	0.3286	0.3067	0.2901	0.2768	0.2659	0.2568	0.2226	0.1820	0.1446	0.1111
10	0.6020	0.4450	0.3733	0.3311	0.3029	0.2823	0.2666	0.2541	0.2439	0.2353	0.2032	0.1655	0.1308	0.1000
12	0.5410	0.3924	0.3264	0.2880	0.2624	0.2439	0.2299	0.2187	0.2098	0.2020	0.1737	0.1403	0.1100	0.0833
15	0.4709	0.3346	0.2758	0.2419	0.2195	0.2034	0.1911	0.1815	0.1736	0.1671	0.1429	0.1144	0.0889	0.0667
20	0.3894	0.2705	0.2205	0.1921	0.1735	0.1602	0.1501	0.1422	0.1357	0.1303	0.1108	0.0879	0.0675	0.0500
24	0.3434	0.2354	0.1907	0.1656	0.1493	0.1374	0.1286	0.1216	0.1160	0.1113	0.0942	0.0743	0.0567	0.0417
30	0.2929	0.1980	0.1593	0.1377	0.1237	0.1137	0.1061	0.1002	0.0958	0.0921	0.0771	0.0604	0.0457	0.0333
40	0.2370	0.1576	0.1259	0.1082	0.0968	0.0887	0.0827	0.0780	0.0745	0.0713	0.0595	0.0462	0.0347	0.0250
60	0.1737	0.1131	0.0895	0.0765	0.0682	0.0623	0.0583	0.0552	0.0520	0.0497	0.0411	0.0316	0.0234	0.0167
120	0.0998	0.0632	0.0495	0.0419	0.0371	0.0337	0.0312	0.0292	0.0279	0.0266	0.0218	0.0165	0.0120	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

PERCENTILE 99

k \ ν	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0.9999	0.9950	0.9794	0.9586	0.9373	0.9172	0.8988	0.8823	0.8674	0.8539	0.7949	0.7067	0.6062	0.5000
3	0.9933	0.9423	0.8831	0.8335	0.7933	0.7606	0.7335	0.7107	0.6912	0.6743	0.6059	0.5153	0.4230	0.3333
4	0.9676	0.8643	0.7814	0.7212	0.6761	0.6410	0.6129	0.5897	0.5702	0.5536	0.4884	0.4057	0.3251	0.2500
5	0.9279	0.7885	0.6957	0.6329	0.5875	0.5531	0.5259	0.5037	0.4854	0.4697	0.4094	0.3351	0.2644	0.2000
6	0.8828	0.7218	0.6258	0.5635	0.5195	0.4866	0.4608	0.4401	0.4229	0.4084	0.3529	0.2858	0.2229	0.1667
7	0.8376	0.6644	0.5685	0.5080	0.4659	0.4347	0.4105	0.3911	0.3751	0.3616	0.3105	0.2494	0.1929	0.1429
8	0.7945	0.6152	0.5209	0.4627	0.4226	0.3932	0.3704	0.3522	0.3373	0.3248	0.2779	0.2214	0.1700	0.1250
9	0.7544	0.5727	0.4810	0.4251	0.3870	0.3592	0.3378	0.3207	0.3067	0.2950	0.2514	0.1992	0.1521	0.1111
10	0.7175	0.5358	0.4469	0.3934	0.3572	0.3308	0.3106	0.2945	0.2813	0.2704	0.2297	0.1811	0.1376	0.1000
12	0.6528	0.4751	0.3919	0.3428	0.3099	0.2861	0.2680	0.2535	0.2419	0.2320	0.1961	0.1535	0.1157	0.0833
15	0.5747	0.4069	0.3317	0.2882	0.2593	0.2386	0.2228	0.2104	0.2002	0.1918	0.1612	0.1251	0.0934	0.0667
20	0.4799	0.3297	0.2654	0.2288	0.2048	0.1877	0.1748	0.1646	0.1567	0.1501	0.1248	0.0960	0.0709	0.0500
24	0.4247	0.2871	0.2295	0.1970	0.1759	0.1608	0.1495	0.1406	0.1338	0.1283	0.1060	0.0810	0.0595	0.0417
30	0.3632	0.2412	0.1913	0.1635	0.1454	0.1327	0.1232	0.1157	0.1100	0.1054	0.0867	0.0658	0.0480	0.0333
40	0.2940	0.1915	0.1508	0.1281	0.1135	0.1033	0.0957	0.0898	0.0853	0.0816	0.0668	0.0503	0.0363	0.0250
60	0.2151	0.1371	0.1069	0.0902	0.0796	0.0722	0.0668	0.0625	0.0594	0.0567	0.0461	0.0344	0.0245	0.0167
120	0.1225	0.0759	0.0585	0.0489	0.0429	0.0387	0.0357	0.0334	0.0316	0.0302	0.0242	0.0178	0.0125	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A-11. Percentage points of the studentized range and coefficients for Tukey's method.

$\alpha = 05$

ν_2 \ K	2	3	4	5	6	7	8	9	10
1	17 97	26 98	32 82	37 08	40 41	43 12	45 40	47 36	49 07
2	6 085	8 331	9 798	10 88	11 74	12 44	13 03	13 54	13 99
3	4 501	5 910	6 825	7 502	8 037	8 478	8 853	9 177	9 462
4	3 927	5 040	5 757	6 287	6 707	7 053	7 347	7 602	7 826
5	3.635	4 602	5 218	5 673	6 033	6 330	6 582	6 802	6 995
6	3 461	4 339	4 896	5 305	5 628	5 895	6 122	6 319	6 493
7	3.344	4 165	4 681	5 060	5 359	5 606	5 815	5 998	6 158
8	3 261	4 041	4 529	4 886	5 167	5 399	5 597	5 767	5 918
9	3 199	3 949	4 415	4 756	5 024	5 244	5 432	5 595	5 739
10	3.151	3 877	4 327	4 654	4 912	5 124	5 305	5 461	5 599
11	3 113	3 820	4 256	4 574	4 823	5 028	5 202	5 353	5 487
12	3 082	3 773	4 199	4 508	4 751	4 950	5 119	5 265	5 395
13	3 055	3 735	4 151	4 453	4 690	4 885	5 049	5 192	5 318
14	3 033	3 702	4 111	4 407	4 639	4 829	4 990	5 131	5 254
15	3 014	3 674	4 076	4 367	4 595	4 782	4 940	5 077	5 198
16	2.998	3 649	4 046	4 333	4 557	4 741	4 897	5 031	5 150
17	2 984	3 628	4 020	4 303	4 524	4 705	4 858	4 991	5 108
18	2.971	3 609	3 997	4 277	4 495	4 673	4 824	4 956	5 071
19	2 960	3 593	3 977	4 253	4 469	4 645	4 794	4 924	5 038
20	2.950	3 578	3 958	4 232	4 445	4 620	4 768	4 896	5 008
24	2 919	3 532	3 901	4 166	4 373	4 541	4 684	4 807	4 915
30	2 888	3 486	3 845	4 102	4 302	4 464	4 602	4 720	4 824
40	2 858	3 442	3 791	4 039	4 232	4 389	4 521	4 635	4 735
60	2 829	3 399	3 737	3 977	4 163	4 314	4 441	4 550	4 646
120	2.800	3 356	3 685	3 917	4 096	4 241	4 363	4 468	4 560
∞	2.772	3 314	3 633	3 858	4 030	4 170	4 286	4 387	4 474

ν_2 \ K	11	12	13	14	15	16	17	18	19
1	50 59	51 96	53 20	54 33	55 36	56 32	57 22	58 04	58 83
2	14 39	14 75	15 08	15 38	15 65	15 91	16 14	16 37	16 57
3	9 717	9 946	10 15	10 35	10 53	10 69	10 84	10 98	11 11
4	8 027	8 208	8 373	8 525	8 664	8 794	8 911	9 028	9 134
5	7.168	7 324	7 466	7 596	7 717	7 828	7 932	8 030	8 122
6	6.649	6 789	6 917	7 034	7 143	7 244	7 338	7 426	7 508
7	6 302	6 431	6 550	6 658	6 759	6 852	6 939	7 020	7 097
8	6 054	6 175	6 287	6 389	6 483	6 571	6 653	6 729	6 802
9	5 867	5 983	6 089	6 186	6 276	6 359	6 437	6 510	6 579
10	5 722	5 833	5 935	6 028	6 114	6 194	6 269	6 339	6 405
11	5 605	5 713	5 811	5 901	5 984	6 062	6 134	6 202	6 265
12	5 511	5 615	5 710	5 798	5 878	5 953	6 023	6 089	6 151
13	5 431	5 533	5 625	5 711	5 789	5 862	5 931	5 995	6 055
14	5 364	5 463	5 554	5 637	5 714	5 786	5 852	5 915	5 974
15	5 306	5 404	5 493	5 574	5 649	5 720	5 785	5 846	5 904
16	5 256	5 352	5 439	5 520	5 593	5 662	5 727	5 786	5 843
17	5 212	5 307	5 392	5 471	5 544	5 612	5 675	5 734	5 790
18	5 174	5 267	5 352	5 429	5 501	5 568	5 630	5 688	5 743
19	5 140	5 231	5 315	5 391	5 462	5 528	5 589	5 647	5 701
20	5.108	5 199	5 282	5 357	5 427	5 493	5 553	5 610	5 663
24	5 012	5 099	5 179	5 251	5 319	5 381	5 439	5 494	5 545
30	4 917	5 001	5 077	5 147	5 211	5 271	5 327	5 379	5 429
40	4.824	4 904	4 977	5 044	5 106	5 163	5 216	5 266	5 313
60	4 732	4 808	4 878	4 942	5 001	5 056	5 107	5 154	5 199
120	4 641	4 714	4 781	4 842	4 898	4 950	4 998	5 044	5 086
∞	4 552	4 622	4 685	4 743	4 796	4 845	4 891	4 934	4 974

H. Leon Harter (Aerospace Research Laboratories, Wright-Patterson AFB), *Annals of Mathematical Statistics* vol. 31, 1960, pp. 1122–1147. Used by permission of author and publisher

Table A-11. Percentage points of the studentized range and coefficients for Tukey's method (*Continued*).

$\alpha = .05$

ν_2 \ K	20	22	24	26	28	30	32	34	36
1	59 56	60 91	62 12	63 22	64 23	65 15	66 01	66 81	67 56
2	16 77	17 13	17 45	17 75	18 02	18 27	18 50	18 72	18 92
3	11 24	11 47	11 68	11 87	12 05	12 21	12 36	12 50	12 63
4	9 233	9 418	9 584	9 736	9 875	10 00	10 12	10 23	10 34
5	8 208	8 368	8 512	8 643	8 764	8 875	8 979	9 075	9 165
6	7 587	7 730	7 861	7 979	8 088	8 189	8 283	8 370	8 452
7	7 170	7 303	7 423	7 533	7 634	7 728	7 814	7 895	7 972
8	6 870	6 995	7 109	7 212	7 307	7 395	7 477	7 554	7 625
9	6 644	6 763	6 871	6 970	7 061	7 145	7 222	7 295	7 363
10	6 467	6 582	6 686	6 781	6 868	6 948	7 023	7 093	7 159
11	6 326	6 436	6 536	6 628	6 712	6 790	6 863	6 930	6 994
12	6 209	6 317	6 414	6 503	6 585	6 660	6 731	6 796	6 858
13	6 112	6 217	6 312	6 398	6 478	6 551	6 620	6 684	6 744
14	6 029	6 132	6 224	6 309	6 387	6 459	6 526	6 588	6 647
15	5 958	6 059	6 149	6 233	6 309	6 379	6 445	6 506	6 564
16	5 897	5 995	6 084	6 166	6 241	6 310	6 374	6 434	6 491
17	5 842	5 940	6 027	6 107	6 181	6 249	6 313	6 372	6 427
18	5 794	5 890	5 977	6 055	6 128	6 195	6 258	6 316	6 371
19	5 752	5 846	5 932	6 009	6 081	6 147	6 209	6 267	6 321
20	5 714	5 807	5 891	5 968	6 039	6 104	6 165	6 222	6 275
24	5 594	5 683	5 764	5 838	5 906	5 968	6 027	6 081	6 132
30	5 475	5 561	5 638	5 709	5 774	5 833	5 889	5 941	5 990
40	5 358	5 439	5 513	5 581	5 642	5 700	5 753	5 803	5 849
60	5 241	5 319	5 389	5 453	5 512	5 566	5 617	5 664	5 708
120	5 126	5 200	5 266	5 327	5 382	5 434	5 481	5 526	5 568
∞	5 012	5 081	5 144	5 201	5 253	5 301	5 346	5 388	5 427

ν \ K	38	40	50	60	70	80	90	100
1	68 26	68 92	71 73	73 97	75 82	77 40	78 77	79 98
2	19 11	19 28	20 05	20 66	21 16	21 59	21 96	22 29
3	12 75	12 87	13 36	13 76	14 08	14 36	14 61	14 82
4	10 44	10 53	10 93	11 24	11 51	11 73	11 92	12 09
5	9 250	9 330	9 674	9 949	10 18	10 38	10 54	10 69
6	8 529	8 601	8 913	9 163	9 370	9 548	9 702	9 839
7	8 043	8 110	8 400	8 632	8 824	8 989	9 133	9 261
8	7 693	7 756	8 029	8 248	8 430	8 586	8 722	8 843
9	7 428	7 488	7 749	7 958	8 132	8 281	8 410	8 526
10	7 220	7 279	7 529	7 730	7 897	8 041	8 166	8 276
11	7 053	7 110	7 352	7 546	7 708	7 847	7 968	8 075
12	6 916	6 970	7 205	7 394	7 552	7 687	7 804	7 909
13	6 800	6 854	7 083	7 267	7 421	7 552	7 667	7 769
14	6 702	6 754	6 979	7 159	7 309	7 438	7 550	7 650
15	6 618	6 669	6 888	7 065	7 212	7 339	7 449	7 546
16	6 544	6 594	6 810	6 984	7 128	7 252	7 360	7 457
17	6 479	6 529	6 741	6 912	7 054	7 176	7 283	7 377
18	6 422	6 471	6 080	6 848	6 989	7 109	7 213	7 307
19	6 371	6 419	6 626	6 792	6.930	7 048	7 152	7 244
20	6 325	6 373	6 576	6 740	6 877	6 994	7 097	7 187
24	6 181	6 226	6 421	6.579	6 710	6 822	6 920	7 008
30	6 037	6 080	6 267	6 417	6 543	6 [illegible]50	6 744	6 827
40	5 893	5 934	6 112	6 255	6 375	6 477	6 566	6 645
60	5 750	5 789	5 958	6 093	6 206	6 303	6 387	6 462
120	5 607	5 644	5 802	5 929	6 035	6 126	6 205	6 275
∞	5 463	5 498	5 646	5 764	5 863	5 947	6 020	6 085

Table A-11. Percentage points of the studentized range and coefficients for Tukey's method (*Continued*).

$\alpha = 01$

ν_2 \ K	2	3	4	5	6	7	8	9	10
1	90 03	135 0	164 3	185 6	202 2	215 8	227 2	237 0	245 6
2	14 04	19 02	22 29	24 72	26 63	28 20	29 53	30 68	31 69
3	8 261	10 62	12 17	13 33	14 24	15 00	15 64	16 20	16 69
4	6 512	8 120	9 173	9 958	10 58	11 10	11 55	11 93	12 27
5	5 702	6 976	7 804	8 421	8 913	9 321	9 669	9 972	10 24
6	5 243	6 331	7 033	7 556	7 973	8 318	8 613	8 869	9 097
7	4 949	5 919	6 543	7 005	7 373	7 679	7 939	8 166	8 368
8	4 746	5 635	6 204	6 625	6 960	7 237	7 474	7 681	7 863
9	4 596	5 428	5 957	6 348	6 658	6 915	7 134	7 325	7 495
10	4 482	5 270	5 769	6 136	6 428	6 669	6 875	7 055	7 213
11	4 392	5 146	5 621	5 970	6 247	6 476	6 672	6 842	6 992
12	4 320	5 046	5 502	5 836	6 101	6 321	6 507	6 670	6 814
13	4 260	4 964	5 404	5 727	5 981	6 192	6 372	6 528	6 667
14	4 210	4 895	5 322	5 634	5 881	6 085	6 258	6 409	6 543
15	4 168	4 836	5 252	5 556	5 796	5 994	6 162	6 309	6 439
16	4 131	4 786	5 192	5 489	5 722	5 915	6 079	6 222	6 349
17	4 099	4 742	5 140	5 430	5 659	5 847	6 007	6 147	6 270
18	4 071	4 703	5 094	5 379	5 603	5 788	5 944	6 081	6 201
19	4 046	4 670	5 054	5 334	5 554	5 735	5 889	6 022	6 141
20	4 024	4 639	5 018	5 294	5 510	5 688	5 839	5 970	6 087
24	3 956	4 546	4 907	5 168	5 374	5 542	5 685	5 809	5 919
30	3 889	4 455	4 799	5 048	5 242	5 401	5 536	5 653	5 756
40	3 825	4 367	4 696	4 931	5 114	5 265	5 392	5 502	5 599
60	3 762	4 282	4 595	4 818	4 991	5 133	5 253	5 356	5 447
120	3 702	4 200	4 497	4 709	4 872	5 005	5 118	5 214	5 299
∞	3 643	4 120	4 403	4 603	4 757	4 882	4 987	5 078	5 157

ν_2 \ K	11	12	13	14	15	16	17	18	19
1	253 2	260 0	266 2	271 8	277 0	281 8	286 3	290 4	294 3
2	32 59	33 40	34 13	34 81	35 43	36 00	36 53	37 03	37 50
3	17 13	17 53	17 89	18 22	18 52	18 81	19 07	19 32	19 55
4	12 57	12 84	13 09	13 32	13 53	13 73	13 91	14 08	14 24
5	10 48	10 70	10 89	11 08	11 24	11 40	11 55	11 68	11 81
6	9 301	9 485	9 653	9 808	9 951	10 08	10 21	10 32	10 43
7	8 548	8 711	8 860	8 997	9 124	9 242	9 353	9 456	9 554
8	8 027	8 176	8 312	8 436	8 552	8 659	8 760	8 854	8 943
9	7 647	7 784	7 910	8 025	8 132	8 232	8 325	8 412	8 495
10	7 356	7 485	7 603	7 712	7 812	7 906	7 993	8 076	8 153
11	7 128	7 250	7 362	7 465	7 560	7 649	7 732	7 809	7 883
12	6 943	7 060	7 167	7 265	7 356	7 441	7 520	7 594	7 665
13	6 791	6 903	7 006	7 101	7 188	7 269	7 345	7 417	7 485
14	6 664	6 772	6 871	6 962	7 047	7 126	7 199	7 268	7 333
15	6 555	6 660	6 757	6 845	6 927	7 003	7 074	7 143	7 204
16	6 462	6 564	6 658	6 744	6 823	6 898	6 967	7 032	7 093
17	6 381	6 480	6 572	6 656	6 734	6 806	6 873	6 937	6 997
18	6 310	6 407	6 497	6 579	6 655	6 725	6 792	6 854	6 912
19	6 247	6 342	6 430	6 510	6 585	6 654	6 719	6 780	6 837
20	6 191	6 285	6 371	6 450	6 523	6 591	6 654	6 714	6 771
24	6 017	6 106	6 186	6 261	6 330	6 394	6 453	6 510	6 563
30	5 849	5 932	6 008	6 078	6 143	6 203	6 259	6 311	6 361
40	5 686	5 764	5 835	5 900	5 961	6 017	6 069	6 119	6 165
60	5 528	5 601	5 667	5 728	5 785	5 837	5 886	5 931	5 974
120	5 375	5 443	5 505	5 562	5 614	5 662	5 708	5 750	5 790
∞	5 227	5 290	5 348	5 400	5 448	5 493	5 535	5.574	5 611

Table A-11. Percentage points of the studentized range and coefficients for Tukey's method (*Continued*).

$\alpha =$ 01

ν_2 \ K	20	22	24	26	28	30	32	34	36
1	298 0	304 7	310 8	316 3	321 3	326 0	330 3	334 3	338 0
2	37 95	38 76	39 49	40 15	40 76	41 32	41 84	42 33	42 78
3	19 77	20 17	20 53	20 86	21 16	21 44	21 70	21 95	22 17
4	14 40	14 68	14 93	15 16	15 37	15 57	15 75	15 92	16 08
5	11 93	12 16	12 36	12 54	12 71	12 87	13 02	13 15	13 28
6	10 54	10 73	10 91	11 06	11 21	11 34	11 47	11 58	11 69
7	9 646	9 815	9 970	10 11	10 24	10 36	10 47	10 58	10 67
8	9 027	9 182	9 322	9 450	9 569	9 678	9 779	9 874	9 964
9	8 573	8 717	8 847	8 966	9 075	9 177	9 271	9 360	9 443
10	8 226	8 361	8 483	8 595	8 698	8 794	8 883	8 966	9 044
11	7 952	8 080	8 196	8 303	8 400	8 491	8 575	8 654	8 728
12	7 731	7 853	7 964	8 066	8 159	8 246	8 327	8 402	8 473
13	7 548	7 665	7 772	7 870	7 960	8 043	8 121	8 193	8 262
14	7 395	7 508	7 611	7 705	7 792	7 873	7 948	8 018	8 084
15	7 264	7 374	7 474	7 566	7 650	7 728	7 800	7 869	7 932
16	7 152	7 258	7 356	7 445	7 527	7 602	7 673	7 739	7 802
17	7 053	7 158	7 253	7 340	7 420	7 493	7 563	7 627	7 687
18	6 968	7 070	7 163	7 247	7 325	7 398	7 465	7 528	7 587
19	6 891	6 992	7 082	7 166	7 242	7 313	7 379	7 440	7 498
20	6 823	6 922	7 011	7 092	7 168	7 237	7 302	7 362	7 419
24	6 612	6 705	6 789	6 865	6 936	7 001	7 062	7 119	7 173
30	6 407	6 494	6 572	6 644	6 710	6 772	6 828	6 881	6 932
40	6 209	6 289	6 362	6 429	6 490	6 547	6 600	6 650	6 697
60	6 015	6 090	6 158	6 220	6 277	6 330	6 378	6 424	6 467
120	5 827	5 897	5 959	6 016	6 069	6 117	6 162	6 204	6 244
∞	5 645	5 709	5 766	5 818	5 866	5 911	5 952	5 990	6 026

ν_2 \ K	38	40	50	60	70	80	90	100
1	341 5	344 8	358 9	370 1	379 4	387 3	394 1	400 1
2	43 21	43 61	45 33	46 70	47 83	48 80	49 64	50 38
3	22 39	22 59	23 45	24 13	24 71	25 19	25 62	25 99
4	16 23	16 37	16.98	17 46	17 86	18 20	18 50	18 77
5	13 40	13 52	14 00	14 39	14 72	14 99	15 23	15 45
6	11 80	11 90	12 31	12 65	12 92	13 16	13 37	13 55
7	10 77	10 85	11 23	11 52	11 77	11 99	12 17	12 34
8	10 05	10 13	10 47	10 75	10 97	11 17	11 34	11 49
9	9 521	9 594	9.912	10 17	10 38	10 57	10 73	10 87
10	9 117	9 187	9 486	9 726	9 927	10 10	10 25	10 39
11	8 798	8 864	9 148	9 377	9.568	9 732	9 875	10 00
12	8 539	8 603	8 875	9 094	9 277	9 434	9 571	9 693
13	8 326	8 387	8 648	9 859	9 035	9 167	9 318	9.436
14	8 146	8 204	8 457	8 661	8.832	8 978	9 106	9 219
15	7 992	8.049	8 295	8 492	8 658	8.800	8 924	9.035
16	7 860	7 916	8 154	8 347	8 507	8 646	8 767	8 874
17	7 745	7 799	8 031	8 219	8 377	8 511	8 630	8 735
18	7 643	7 696	7 924	8 107	8 261	8 393	8 508	8 611
19	7 553	7 605	7 828	8 008	8 159	8 288	8 401	8 502
20	7 473	7 523	7 742	7 919	8 067	8 194	8 305	8.404
24	7 223	7 270	7 476	7 642	7 780	7 900	8 004	8 097
30	6 978	7 023	7 215	7 370	7 500	7 611	7 709	7 796
40	6 740	6 782	6 960	7 104	7 225	7 328	7 419	7 500
60	6 507	6 546	6 710	6 843	6 954	7 050	7 133	7 207
120	6 281	6 316	6 467	6 588	6 689	6 776	6 852	6 919
∞	6 060	6 092	6 228	6 338	6 429	6 507	6 575	6.636

Table A-12. Fisher's Z transformation.

r	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.00000	01000	02000	03001	04002	.05004	.06007	07012	08017	09024
.1	10034	11045	12058	13074	14093	15114	16139	17167	18198	.19234
2	.20273	21317	22366	23419	24477	25541	.26611	27686	28768	.29857
3	30952	32055	33165	34283	35409	.36544	37689	38842	40006	41180
.4	.42365	43561	44769	45990	47223	.48470	49731	51007	.52298	.53606
.5	.54931	56273	57634	59014	60415	.61838	63283	64752	66246	67767
6	.69315	70892	.72500	74142	.75817	.77530	.79281	81074	82911	84795
.7	.86730	88718	90764	92873	95048	.97295	99621	1 02033	1 04537	1 07143
8	1 09861	1 12703	1 15682	1 18813	1 22117	1.25615	1 29334	1 33308	1 37577	1.42192
9	1 47222	1.52752	1 58902	1 65839	1 73805	1.83178	1 94591	2 09229	2 29756	2.64665

For negative values of r put a minus sign in front of the tabled numbers.

Table A-13. Coefficients for orthogonal polynomials for trend analysis.

k	Polynomial	$X = 1$	2	3	4	5	6	7	8	9	10	$\Sigma\xi'^2$	λ
3	Linear	−1	0	1								2	1
	Quadratic	1	−2	1								6	3
	Linear	−3	−1	1	3							20	2
4	Quadratic	1	−1	−1	1							4	1
	Cubic	−1	3	−3	1							20	$\frac{10}{3}$
	Linear	−2	−1	0	1	2						10	1
5	Quadratic	2	−1	−2	−1	2						14	1
	Cubic	−1	2	0	−2	1						10	$\frac{5}{6}$
	Quartic	1	−4	6	−4	1						70	$\frac{35}{12}$
	Linear	−5	−3	−1	1	3	5					70	2
6	Quadratic	5	−1	−4	−4	−1	5					84	$\frac{3}{2}$
	Cubic	−5	7	4	−4	−7	5					180	$\frac{5}{3}$
	Quartic	1	−3	2	2	−3	1					28	$\frac{7}{12}$
	Linear	−3	−2	−1	0	1	2	3				28	1
7	Quadratic	5	0	−3	−4	−3	0	5				84	1
	Cubic	−1	1	1	0	−1	−1	1				6	$\frac{1}{6}$
	Quartic	3	−7	1	6	1	−7	3				154	$\frac{7}{12}$
	Linear	−7	−5	−3	−1	1	3	5	7			168	2
	Quadratic	7	1	−3	−5	−5	−3	1	7			168	1
8	Cubic	−7	5	7	3	−3	−7	−5	7			264	$\frac{2}{3}$
	Quartic	7	−13	−3	9	9	−3	−13	7			616	$\frac{7}{12}$
	Quintic	−7	23	−17	−15	15	17	−23	7			2184	$\frac{7}{10}$
	Linear	−4	−3	−2	−1	0	1	2	3	4		60	1
	Quadratic	28	7	−8	−17	−20	−17	−8	7	28		2772	3
9	Cubic	−14	7	13	9	0	−9	−13	−7	14		990	$\frac{5}{6}$
	Quartic	14	−21	−11	9	18	9	−11	−21	14		2002	$\frac{7}{12}$
	Quintic	−4	11	−4	−9	0	9	4	−11	4		468	$\frac{3}{20}$
	Linear	−9	−7	−5	−3	−1	1	3	5	7	9	330	2
	Quadratic	6	2	−1	−3	−4	−4	−3	−1	2	6	132	$\frac{1}{2}$
10	Cubic	−42	14	35	31	12	−12	−31	−35	−14	42	8580	$\frac{5}{3}$
	Quartic	18	−22	−17	3	18	18	3	−17	−22	18	2860	$\frac{5}{12}$
	Quintic	−6	14	−1	−11	−6	6	11	1	−14	6	780	$\frac{1}{10}$

B LIST OF STATISTICAL TERMS

Acceptance region In hypothesis testing, the acceptance region consists of the outcomes in the sample space or of a statistic based upon the outcomes that do not lead to a rejection of a tested hypothesis.

Accuracy In the estimation of a parameter, accuracy refers to the closeness that exists between an estimated value and the true unknown value of the parameter being estimated.

Alternative hypothesis In hypothesis testing, an alternative hypothesis refers to any hypothesis that is alternative to the one under test. It is usually denoted by H_1 and is associated with the research hypothesis of an investigation.

Arithmetic mean The arithmetic mean of a set of numerical values is their sum divided by their number.

Average Average refers to a measure of central tendency such as the arithmetic mean, the sample median, or the mode. In most situations it is a synonym for the arithmetic mean.

Bernoulli variable A variable that can assume only one of two possible values in a given trial is called a Bernoulli variable, if, in addition, the probabilities of the two possible values or states remain constant at p, and $q = 1 - p$.

Bernoulli trials A sequence of trials that are independent and such that each trial gives rise to a Bernoulli variable is called a Bernoulli set of trials Their associated probabilities are given by the binomial formula

$$P(X=x)=\binom{N}{x}p^{x}q^{n-x}$$

Bimodal distribution A probability or relative frequency distribution that has two modes is said to be bimodal.

Bivariate distribution The joint probability distribution of two variables that vary together is said to be a bivariate probability distribution

Cartesian product If A and B are sets, then the set of ordered pairs generated by elements of the form (a,b) are said to constitute the cartesian product of the sets A and B. In generating this set, every element of A is associated with every element of B.

Centroid The centroid refers to the exact center of a bivariate distribution

Combination A combination is a set of objects. The number of different combinations of R objects that can be generated from N objects is given by

$$\binom{N}{R}=\frac{N!}{R!(N-R)!}$$

Complementary set If A is a subset of S, then all of the remaining elements of S that are not in A define the complementary set of A.

Composite hypothesis A composite hypothesis consists of a set of simple hypotheses. The hypothesis $H_0: \mu = 15$ is simple, while the hypothesis $H_0: \mu_1 = \mu_2$ is composite since any real number can be assumed by the parameters.

Confidence interval A confidence interval is a set of numerical values that could conceivably be exact values of an unknown parameter. The interval $\theta_1 < \theta < \theta_2$ states that the parameter θ is larger than θ_1 but less than θ_2.

Continuous variable A variable is said to be continuous if it can assume any of the values in a continuous range of numerical values.

Contrast A contrast is a weighted sum of parameters such that the sum of the weights add to zero.

Correlation Correlation refers to the association that exists between two variables. It also refers to the tendency that two variables have to vary together

Covariate A variable that is used to help explain some of the variability in a dependent measure is called a covariate. Covariates are generally observed in an analysis-of-covariance design to adjust the values of the dependent variable.

Critical range, or rejection region In hypothesis testing, the critical region consists of the outcomes in the sample space that lead to a rejection of a tested hypothesis.

Critical value The critical value is the value of a test statistic that partitions its range of values into a rejection and an acceptance region for the testing of a statistical hypothesis.

Curvilinear A regression curve that is not linear is said to be curvilinear.

Decile A decile refers to one of the nine numerical values that partition a probability distribution into ten equal parts.

Decision rule In the theory of hypothesis testing, a decision rule is a rule that specifies when a statistical hypothesis under test should be rejected.

Dependent variable In a regression model, the dependent variable is the variable that is observed. In an analysis-of-variance model, the dependent variable is frequently called the criterion variable.

Derived statistic A statistic that is determined from other data by a sequence of arithmetic operations is called a derived statistic. The original statistics are frequently called raw data or raw statistics.

Discrete variable A variable that can take on only a set of discrete, unique numerical values is said to be a discrete variable.

Disjoint or mutually exclusive sets If two sets A and B have no members in common, they are said to be disjoint or mutually exclusive sets.

Efficiency In the theory of parameter estimation, an estimator is regarded as being more efficient than another competing estimator if the variance of the sampling distribution of the estimator is smaller than the variance of the competing estimator.

Elements of a set The objects that constitute a set are called its elements.

Equal sets Two sets A and B are said to be equal if they contain the same exact elements.

Event The result of an investigation or experiment is called an event. For example, the random toss of a coin gives rise to one of the two events {heads, tails}

Expected value The expected value of a statistic is its long-term average over repeated samplings.

Experiment An experiment is a random selection of an event from an appropriate sample space.

Finite population A population that has a finite number of members is said to be a finite population.

Homoscedasticity Homoscedasticity refers to the assumption in regression theory that states that the variances of all conditional distributions are identical.

Hypothesis A statistical hypothesis is a statement regarding the parameters or form of a probability distribution that is eventually put to a statistical test.

Independence Two characteristics A and B are said to be statistically independent if $P(A \cap B) = P(A)P(B)$.

Independent variable In regression theory, an independent variable X refers to the variable that is used to predict the value of the criterion measure Y. It is also used to represent the manipulated variable in a controlled study or experiment.

Median The median value of a distribution is the value that divides a distribution in half.

Mode The mode of a distribution is the value with greatest frequency, probability, or likelihood of appearance.

Monotonic The relationship between two variables is said to be monotonic if as one variable increases in numerical value, the other always increases or always decreases.

One-sided test A test of hypothesis in which the rejection region is placed entirely at one end of the sampling distribution of the test statistic is called a one-tailed test. If the rejection region is placed in both tails, it is called a two-tailed test of hypothesis.

Outlier An outlier is a sample value that appears as though it did not come from the population from which the rest of the sample was generated.

Parameter A parameter is a fixed numerical quantity that characterizes a probability distribution.

Percentile A percentile refers to one of the 99 values that partition a probability distribution into 100 equal parts.

Permutation A permutation is an ordering of N objects. If the objects are all distinct, then the number of different permutations or orderings is given by $N! = 1 \times 2 \times 3 \times \cdots \times N$.

Point estimate A point estimate is a single numerical value that is used to characterize an unknown parameter of a probability distribution. An interval estimate consists of a set of point estimates and is generally called a confidence interval.

Population, or universe A population is any collection of objects, real or imagined, that can serve as the subject of a research investigation.

Power In hypothesis testing, power refers to the probability of rejecting a false hypothesis that is under test.

Precision Precision refers to the tendency that a sampling distribution of a statistic has to cluster about the value being estimated. A very precise estimator tends to cluster in a very compact manner about the value being estimated.

Probability For a finite universe with $N(S)$ elements, consider the elements that have the property A. Let the number of such elements be $n(A)$. If all elements of the universe are equally likely, then the ratio of $n(A)$ to $N(S)$ is called the probability of A. For a nonfinite universe, the probability of A is given by the limit of the ratio $n(A)$ to N where N refers to the number of trials that is allowed to increase without limit.

Probability distribution A probability distribution consists of a variable and the probabilities of occurrence of the various values of the variable.

Proper subset B is said to be a proper subset of A if all elements of B are also elements of A.

Qualitative variable A qualitative variable is one in which a characteristic is specified by class exclusion or inclusion. For example, sex is a qualitative variable for which the possible classes are male and female.

Quartile A quartile refers to one of three numerical values that partition a probability distribution into four equal parts.

Random, or chance Random is an undefined concept in probability theory and in this book It is an adjective that is generally associated with a special type of sampling called random sampling.

Random sampling Random sampling refers to a method of sampling in which all elements of the population have an equal chance of being included in the sample and for which sampling is statistically independent

Random variable A random variable refers to a characteristic that may vary over a specified range of numerical values with a certain probability distribution The range of values may be finite or infinite, discrete or continuous.

Range The difference between the largest and smallest element of a sample or population is called the range of the sample or the population.

Regression equation A regression equation defines the expected values of the conditional distributions of a variable Y in terms of another variable X. Y is called the dependent variable, while X is called the independent variable.

Relative frequency The relative frequency of an event is the ratio of the number of times the event occurs over N trials to the number of trials. If the trials are allowed to increase without limit, then the relative frequency converges on the probability of the event.

Residual variance or mean square within, or mean square residual, or unexplained variance The residual variance is the part of the variance that remains in a set of data after all explanatory sources of variance have been removed.

Sample A sample is a subset of a population or universe

Sample space A sample space consists of all possible outcomes or samples that could be observed in an experiment or investigation.

Sampling distribution If one computes the same statistic over the complete totality of possible samples, the probability distribution so generated is called the sampling distribution of the statistic.

Scatter diagram A scatter diagram is a graphic presentation of the joint distribution of two variables measured on the elements in a sample.

Set A set is a well-defined collection of distinct objects of any sort. To say that a set is well defined is to say that it is always possible to determine whether or not a specific object is, or is not, a member of the set.

Set intersection If A and B are subsets of S, then the elements that are common to both A and B constitute the set that is called the intersection of A and B.

Set union If A and B are subsets of S, then the set of elements consisting of A or B or their intersection constitutes the set that is called the union of A and B.

Simple hypothesis A statistical hypothesis that completely specifies the parameters of a probability distribution is said to be a simple hypothesis.

Skewed distribution A distribution that is not symmetrical about its center is said to be skewed.

Slope The slope of a regression equation refers to the constant that specifies the magnitude of the increase in the dependent variable for each unit increase in the independent variable.

Standard deviation The standard deviation of a probability distribution is a measure of its variability relative to its center. In a sample, the standard deviation is defined by the following formula:

$$S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{N - 1}}$$

Standard error A standard error of a statistic is an estimate of the standard deviation of the sampling distribution of the statistic.

Standard score A standard score is a transformed score that indicates the number of standard deviations between an observed value and its expected or mean value. For a sample value and for a population value X, the associated standard score is given by $Z = (X - \bar{X})/S$ and $Z = (X - \mu)/\sigma$.

Statistic A statistic is a summary measure computed on observed sample values.

Statistical test A statistical test is made whenever a statistical hypothesis is subjected to a rule that states when the hypothesis should or should not be rejected.

Symmetrical distribution If the probabilities of values equidistant from the center of a distribution are equal, it is said that the distribution is symmetrical about its center.

Type I error A type I error is the incorrect rejection of a true hypothesis.

Type II error A type II error is the nonrejection of a false hypothesis.

Unbiased estimator If the expected value of the sampling distribution of a statistic used as an estimator is exactly equal to the parameter being estimated, it is said that the estimator is unbiased.

Variable A variable is a characteristic that varies across the elements of a population or sample.

C ANSWER KEY

CHAPTER 1

1-7 Qualitative (*e*), (*g*), (*m*).
Discrete (*b*), (*c*), (*d*), (*f*), (*i*), (*k*), (*l*), (*p*), (*r*), (*t*)
Continuous (*a*), (*h*), (*j*), (*n*), (*o*), (*q*), (*s*).

CHAPTER 2

2-2 Finite (*a*), (*b*), (*d*), (*f*), (*g*), (*h*), (*i*).
Countable infinite (*c*).
Uncountable infinite (*e*), (*j*).

2-5 (*a*) 87; (*b*) 18; (*c*) 87; (*d*) 73; (*e*) 24.

2-6 {(control, male), (treatment, male), (control, female), (treatment, female)}.

CHAPTER 3

3-2 (*a*) $\frac{1}{36}$; (*b*) $\frac{1}{6}$; (*c*) $\frac{1}{6}$; (*d*) $\frac{11}{36}$; (*e*) $\frac{5}{36}$; (*f*) $\frac{5}{36}$; (*g*) 0; (*h*) $\frac{10}{36}$.

3-5 (*a*) $\frac{41}{52}$; (*b*) $\frac{11}{52}$; (*c*) $\frac{12}{23}$; (*d*) $\frac{22}{29}$; (*e*) $\frac{34}{52}$; (*f*) No; (*g*) $\frac{18}{52}$; (*h*) 1; (*i*) No

3-6 (*a*) .4; (*b*) .9; (*c*) 0, (*d*) 0; (*e*) $\frac{6}{7}$; (*f*) .1; (*g*) .75; (*h*) 0; (*i*) .1; (*j*) 1.

3-7 (*a*) .4; (*b*) .72; (*c*) .18; (*d*) .6; (*e*) .6; (*f*) .28; (*g*) .3; (*h*) .3; (*i*) .28; (*j*) .82.

3-9 (*a*) $\frac{87}{146}$; (*b*) $\frac{18}{146}$; (*c*) $\frac{87}{146}$; (*d*) $\frac{73}{146}$; (*e*) $\frac{24}{146}$.

3-10 (*a*) .30; (*b*) .54; (*c*) .44; (*d*) .21.

CHAPTER 4

4-1 6 branches.

4-2 19 branches.

4-4 10!

4-5 $\binom{10}{3} = \binom{10}{7} = \frac{10!}{3!7!}$

4-6 $\binom{15}{6}$ or $\binom{15}{9}$

4-7

Y	Probability
0	$\dfrac{\binom{20}{10}\binom{5}{0}}{\binom{25}{10}} = .05652$
1	$\dfrac{\binom{20}{9}\binom{5}{1}}{\binom{25}{10}} = .25691$
2	$\dfrac{\binom{20}{8}\binom{5}{2}}{\binom{25}{10}} = .38537$
3	$\dfrac{\binom{20}{7}\binom{5}{3}}{\binom{25}{10}} = .23715$
4	$\dfrac{\binom{20}{6}\binom{5}{4}}{\binom{25}{10}} = .05928$
5	$\dfrac{\binom{20}{5}\binom{5}{5}}{\binom{25}{10}} = .00474$

4-9 (*a*) $\dfrac{\dbinom{30}{2}\dbinom{220}{4}}{\dbinom{250}{6}}$

(*b*) $\dfrac{\dbinom{20}{2}\dbinom{230}{4}}{\dbinom{250}{6}}$

(*c*) $\dfrac{\dbinom{20}{1}\dbinom{10}{1}\dbinom{220}{4}}{\dbinom{250}{6}}$

(*d*) $\dfrac{\dbinom{30}{0}\dbinom{220}{6}}{\dbinom{250}{6}}$

(*e*) $\dfrac{\dbinom{30}{2}\dbinom{220}{4}}{\dbinom{250}{6}} \times \dfrac{\dbinom{2}{1}\dbinom{4}{2}}{\dbinom{6}{3}}$

4-10 (*a*) 10; (*b*) 11.6; (*c*) 25.2; (*d*) –.10; (*e*) 1.6; (*f*) 887; (*g*) 150.8; (*h*) 370.4; (*i*) 542; (*j*) 78.

CHAPTER 5

5-1 (*a*) $\dbinom{6}{0}(.8)^0(.2)^6 = .000064$; (*b*) .65536; (*c*) .34455; (*d*) .90112; (*e*) .26214.

5-2 (*a*) 0; (*b*) 79; (*c*) 41; (*d*) 108; (*e*) 31.

5-3 $P(X = x) = \dbinom{7}{x}(.5)^x(.5)^{7-x}$

5-4 $P(X = x) = \dbinom{7}{x}(.9)^x(.1)^{7-x}$

5-10 (*a*) .001231; (*b*) .16777; (*c*) .01041; (*d*) .06775.

CHAPTER 6

6-1 $\hat{M} = 3.5$, $\hat{P}_{80} = 6$.

6-2 Mode = 5, $\hat{M} = 3.5$, $E(X) = 3.9$.

6-3 $\sigma^2 = 7.23$, $\sigma = \sqrt{7.23} = 2.69$

6-4 $E(T) = 5(3.9) = 19.5$, $\mathrm{Var}(T) = 5(7.23) = 36\ 15$, $\sigma_T = 6.01$.

6-5

X	$P(X = x)$
0	.2058
1	.4018
2	.2898
3	.0918
4	.0108

$E(X) = 1.3$, $\sigma_X = .93$, teacher should be surprised.

6-7

	Expected value	*Variance*	*Standard deviation*
Week 1	4.8	3.36	1.83
Week 2	7.2	5 04	2 24
Week 3	9.9	6.93	2.63
Week 4	8.1	5.67	2.38
Month	30	21	4.58

6-8 Table 5-9 : $E(X) = 1.25$, $\sigma_X = .968$.
Table 5-10 : $E(X) = 1.25$, $\sigma_X = .84$.

6-10

School	*Expected value*	*Standard deviation*
A	18.1	2.10
B	9.6	2.13

CHAPTER 7

7-1 (*a*) .1711; (*b*) .9906.
7-2 (*a*) .5634; (*b*) .2119; (*c*) .0410.
7-3 68 percent range; $13.49 < X < 25.51$; $P(T \geqslant 25) = .203$.
7-4 (*a*) .0071; (*b*) 12.54; (*c*) $10.432 < X < 13.568$.
7-5 (*a*) .0000; (*b*) 24.87; (*c*) $21.79 < X < 26.21$.
7-6 (*a*) .16; (*b*) $Q_1 = \$139.84$, $Q_3 = \$156.16$.
7-7 Hire Miss Black.
7-8 (*a*) 413; (*b*) 576.
7-9 (*a*) 75; (*b*) 464.
7-10 (*a*) .0256; (*b*) $50.40 < X < 69.60$.

CHAPTER 8

8-5 (c) $\bar{X}_C = 13.8$, $\bar{X}_V = 22.07$; $S_C = 5\ 11$, $S_V = 5.27$.

8-6 Tract

0001-A $\hat{P}_{10} = 3.5$, $\hat{P}_{50} = 9.8$, $\hat{P}_{90} = 14.7$.

0006 $\hat{P}_{10} = 12.0$, $\hat{P}_{50}$ = greater than four years of college, $\hat{P}_{90}$ = greater than four years of college.

8-7

	Completed semester 1	2	3	4	5
$\bar{X}$	\$3111	\$4310	\$6153	\$6545	\$7890

8-8 (a)

	With X = 63	*Without X = 63*
$\bar{X}$	24.42	20.91
S D.	14.72	8.72

(b) $X = 63$ is an outlier· $T = 2.6$.

8-10 (a) $\bar{X} = .875$, $S_X = 1.044$.

CHAPTER 9

9-4 $95.55 < \bar{X} < 104.45$.

9-5 $18.07 < Y < 31.93$.

9-6 $\bar{X} > 96.28$, $Y > 30.8$.

9-7 $N = 40$.

9-8 $N = 97$.

CHAPTER 10

10-1 (a) .0717; (b) 1.240; (c) 3.247; (d) 5.892; (e) 22.307; (f) 31.41; (g) 59.34; (h) 88.38; (i) 163.64.

10-2 (a) 133.945; (b) 331.647; (c) 489.467.

10-3 (a) −1.533; (b) −.713; (c) 0; (d) .697; (e) 1.714; (f) 2.045; (g) 2.415; (h) 2.638.

10-4 $\hat{P}_{25} = .1015$, $\hat{P}_{75} = 1.3233$.

10-7 (a) $10.8 < S^2 < 76.08$; (b) $16\ 89 < S^2 < 62.11$; (c) $21.81 < S^2 < 53.6$.

10-8 (a) $20.7 < \bar{X} < 29.3$; (b) $22.2 < \bar{X} < 27.8$; (c) $23.08 < \bar{X} < 26.92$.

CHAPTER 11

11-1 (a) $\bar{X} = 1.235$, $S^2 = 1.566$; (b) $.592 < \mu < 1.878$, $.8686 < \sigma^2 < 3.627$.

11-2 (*a*)

	Conditions None	1	5
$\bar{X}$	45.78	48.88	56.33
S	50 69	45.81	59.75

11-3 $.046 < p_{C|M} - p_{C|F} < .221$.
11-4 $.161 < p_A - p_B < .279$.
11-5 $36.17 < \mu < 37.83$.
11-6 $3.82 < \mu_V - \mu_C < 12.67$.

11-7 (*a*)

	Club members	*Fathers*
$\bar{X}$	2.37	3.73
S	1.69	1.61

(*b*) $.83 < \mu_F - \mu_{CM} < 1.889$.
11-8 (*a*) $-1.511 < \mu_d < 7.329$; (*c*) $21.02 < \sigma_d^2 < 133.0$.
11-9 (*a*) $.83 < \mu_C - \mu_E < 9.37$; (*c*) $38.791 < \sigma_C^2 < 154.829$, $4.36 < \sigma_E^2 < 14.60$.
11-10 (*a*) $200.1 < \mu < 255.9$; (*c*) $17.65 < \sigma_1^2 < 70.52$, $10.19 < \sigma_2^2 < 37.62$.

CHAPTER 12

12-2 $Z = -4.37$; reject H_0; $p \neq .3$.
12-3 H_0: $\mu = 3.90$; $Z = 2.98$; reject H_0; $\mu \neq 3.90$.
12-7 H_0: $\mu = 100$; $Z = -4.75$; reject H_0; $\mu \neq 100$.
12-8 (*d*) reject H_0 if $Z > 1.645$; (*e*) $Z = 3.37$, reject H_0; $p > \frac{1}{2}$.

CHAPTER 13

13-1 H_0: $\mu_1 = \mu_2$, $t = 1.16$, do not reject H_0.
13-2 H_0: $p_A = p_B$, $Z = 4.30$, reject H_0.
13-3 H_0: $\mu_1 = \mu_2$, $t = -4.96$, reject H_0.
13-4 H_0: $\mu_C = \mu_V$, $t = 3.86$, reject H_0.
13-5 H_0: $\mu_d = 0$, $t = 1.52$, do not reject H_0.
13-6 H_0: $\mu_C = \mu_E$, $t^* = 2.5$, $\nu^* = 19$, reject H_0.
13-7 H_0: $p_{CA} = p_{NA}$, $Z = .50$, do not reject; H_0: $p_{CB} = p_{NB}$, $Z = 2.13$, reject.
13-8 H_0: $\mu_d = 0$, $t = -5.27$, reject H_0.

CHAPTER 14

14-1 (*a*) .077; (*b*) .234; (*c*) .248; (*d*) .490; (*e*) 1.55; (*f*) 1.43; (*g*) 1.48; (*h*) 8.10.
14-2 (*a*) 1.12; (*b*) 1.67; (*c*) 1.29; (*d*) 1.07; (*e*) 1.04; (*f*) 1.02; (*g*) 1; (*h*) 1.11.
14-3 H_0: $\sigma_m^2 = \sigma_{\bar{m}}^2$; $F = S_{\bar{m}}^2/S_m^2 = 1.22$; do not reject.
14-4 H_0: $\sigma_M^2 = \sigma_F^2$; $F = S_M^2/S_F^2 = 2.34$; reject H_0.
14-5 Do not reject.

14-7 $H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_5^2$; $\chi^2 = 8.92$; $\chi_4^2(.95) = 9.45$; do not reject.
14-8 $.151 < \sigma_1^2/\sigma_0^2 < 6.94$.
$.207 < \sigma_5^2/\sigma_0^2 < 6.77$.
14-9 $.649 < \sigma_S^2/\sigma_F^2 < 1.81$.
14-10 $3.43 < \sigma_C^2/\sigma_E^2 < 24.30$.

CHAPTER 15

15-1 $\bar{X}_{VW} = 40.2$; $\bar{X}_{FW} = 37.4$; $\bar{X}_{NVW} = 26.7$; $\bar{X}_{NAA} = 23.0$; $F = \text{MSB}/\text{MSW} = \frac{375.3}{91.24} = 4.11$; $F_{3,21}(.95) = 3.07$; reject H_0.
15-3 $\bar{X}_{SA} = 2.94$; $\bar{X}_{MA} = 3.09$; $\bar{X}_{MD} = 3.07$; $\bar{X}_{SD} = 3.00$; $F = \text{MSB}/\text{MSW} = \frac{.467}{.974} = .48$; do not reject H_0.
15-4 $F = \text{MSB}/\text{MSW} = \frac{925.00}{34.38} = 26.9$; reject H_0.
15-6 $\text{I} = 2083$; $\text{II} = .3750$; $\text{III}_C = 1155.6$.
15-7 $F = 14.92$; $t = 3.87$; $(3.87)^2 = 14.98$.
15-8 $F = 9.29$; reject H_0.
15-9 $F = \text{MSB}/\text{MSW} = \frac{262.5}{52.7} = 4.98$; reject.
15-10 $F = \text{MSB}/\text{MSW} = \frac{217.2}{52.9} = 4.11$; reject.

CHAPTER 16

16-1 $X^2 = 35.04$; reject H_0.
16-2 $X^2 = 1.30$; do not reject.
16-3 $X^2 = 8.24$; $\nu = 5$; do not reject.
16-4 $X^2 = 6.54$, $\nu = 6$; do not reject.
16-5 $X^2 = 35.47$; $\nu = 6$; reject H_0.
16-6 $X^2 = 4.53$; $\nu = 3$; do not reject.
16-7 $X^2 = 30.81$; $\nu = 6$; reject H_0.
16-9 $X^2 = 16.78$; $\nu = 6$; reject H_0.

CHAPTER 17

17-1 $\hat{\phi}_1 = .534$; $\hat{\phi}_2 = .388$.
17-5 $X^2 = 2.68$; $\nu = 6$; do not reject.
17-6 Test of independence; $X^2 = 33.27$; $\hat{\phi} = .587$; reject H_0.

CHAPTER 18

18-4 $r = .07$.
18-7 $r = .3389$; $-.076 < E(Z) < .782$; no, $\rho = 0$ is possible.
18-8 $r = .452$.

CHAPTER 19

19-5 $Q = 30.27$; $\nu = 3$; reject H_0.
19-6 (*a*) $S_d^2 = 40.53$; $r = .452$. (*b*) $F = 1.15$; do not reject.
19-7 (*c*) $\hat{Y}_1 = 2.225 + 1.1\,Y_2$; $\hat{Y}_2 = 2.809 + .764\,Y_1$.
(*d*) $E(Y_1 | Y_2 = 60) = 68.225$; $E(Y_2 | Y_1 = 60) = 48.649$.
(*g*) $r_S = .747$; $r = .917$.
19-8 $H_0: \rho = .6$; $Z = .2127$; do not reject.
19-9 $H_0: \rho_1 = \rho_2 = \rho_3 = \rho_4$; $U_0 = 2.179$; $\chi_3^2(.95) = 7.82$; do not reject.

CHAPTER 20

20-4 **Analysis-of-variance table.**

Source	*d/f*	*S of S*	*MS*	*F*
Between groups	2	527		
Linear	1	521.5	521.5	9.90
Curvilinear	1	4.5	4.5	.09
Within	22	1159	52.7	
Total	24	1685		

20-5 $\hat{Y}_X = 13.13 + .543X$; $H_0: \beta = 0$; $t = 2.435$; $\nu = 9$; reject H_0.

20-8 **Analysis-of-variance table.**

Source	*d/f*	*S of S*	*MS*	*F*
Between	3	1155		
Linear	1	1030.4	1030 4	22 2
Curvilinear	2	124.6	62 3	1.3
Within	20	928	46.4	
Total	23	2083		

20-9 **Analysis-of-variance table.**

Source	*d/f*	*S of S*	*MS*	*F*
Between trials	4	79.65		
Linear	1	77 95	77.95	46.67
Curvilinear	3	1.70	.57	.34
Between subjects	7	109.00	27.14	16.25
Within	28	46.75	1.67	
Total	39	316.40		

20-10 $4.68 < \psi_{\text{linear}} < 15.09$; $-7.26 < \psi_{\text{quad}} < 6.00$.

D SUPPLEMENTARY READINGS IN RELATED TEXTS

BLALOCK, H. M. *Social Statistics*, New York: McGraw-Hill, 1960.

DIXON, W., and F. MASSEY. *Introduction to Statistical Analysis*, 3d ed., New York: McGraw-Hill, 1969.

FERGUSON, G. A. *Statistical Analysis in Psychology and Education*, New York: McGraw-Hill, 1959.

GUILFORD, J. P. *Fundamental Statistics in Psychology and Education*, 2d ed., New York: McGraw-Hill, 1965.

GUENTHER, W. C. *Analysis of Variance*, Englewood Cliffs, N.J.: Prentice-Hall, 1964.

GOLDBERG, S. *Probability, An Introduction*, Englewood Cliffs, N.J.: Prentice-Hall, 1960.

HADLEY, G. *Elementary Statistics*, San Francisco: Holden-Day, 1969.

HAYS, W. L. *Statistics for Psychologists*, New York: Holt, Rinehart, and Winston, 1963.

HODGES, J. L., JR., and E. L. LEHMANN. *Basic Concepts of Probability and Statistics*, San Francisco: Holden-Day, 1964.

KERLINGER, F. *Foundations of Behavioral Research*, New York: Holt, Rinehart, and Winston, 1964.

LI, C. C. *Introduction to Experimental Statistics*, New York: McGraw-Hill, 1964.

MENDENHALL, W. *Introduction to Probability and Statistics*, 2d ed., Belmont, California: Wadsworth, 1967.

WALLIS, W. A., and H. V. ROBERTS. *Statistics, a New Approach*, New York: The Free Press of Glencoe, Inc., 1956

WYATT, W. W., and C. M. BRIDGES, JR. *Statistics for the Behavioral Sciences*, Boston: D. C. Heath, 1967.

Table of supplementary readings according to chapter numbers.

Author	*Chapter number*											
MARASCUILO	1	2	3	4	5	6	7	8	9	10	11	12
Blalock	1		9		10		7	4, 5, 6		11, 15	12	8, 11
Dixon–Massey	1, 4		20				5	2, 3	4, 9		6	6, 7, 14
Ferguson	1		5		5	5	6	2, 3, 4	9			10
Guilford	1		7		7	7	7, 19	2, 3, 4, 5	8			8, 10
Guenther												1
Goldberg		1, 2	2	3	5	4, 5	4					
Hadley		1	1		2	2	2		3	3	3	4
Hays	0	1	2, 4	5	5	3	8	3, 6	7, 8	10, 11	9, 10	9, 10
Hodges–Lehmann		1	1, 2, 4	2	3	5, 6	6		8, 9, 10			11, 13
Kerlinger	1, 2, 3	5	8					7	4			9, 10
Li												
Mendenhall	1		4		6	5	7	3	8	9	8, 9	8, 9
Wallis–Roberts	1, 4		10				11	6, 7, 8	2, 11, 14			12
Wyatt–Bridges	1		4	4	4		5	2, 3		5		6, 7

Table of supplementary readings according to chapter numbers (*Continued*).

Author	*Chapter number*							
MARASCUILO	13	14	15	16	17	18	19	20
Blalock	13		16	15	15	17	18	20
Dixon–Massey	8	8	10, 15			11		12
Ferguson			15	11		7	8, 12, 13	
Guilford	9	9	13		11, 14	6, 14	15	15
Guenther	1	1	2					6
Goldberg						4		
Hadley	4		6			7	7	7
Hays	10	11	12, 14	17	17	15	15	16
Hodges–Lehmann	12					7		
Kerlinger			11				13	
Li			5, 6, 7, 8, 31, 32			13		14, 30
Mendenhall	9		12		11	10		
Wallis–Roberts	13					9, 17		
Wyatt–Bridges	7	5, 8	10	9		11	12, 15	14

INDEX

Adjusted means, 505–508
Analysis:
of covariance, 497–508
of variance, 349–356, 458–464, 476–479, 494–497

Bartlett's test for variances, 341, 346–347
Bernoulli variables, 86, 92, 131, 223, 293
Binomial coefficient, 74, 76–78
Binomial distribution, 112, 115, 128–129, 131, 135, 162, 167–168, 223–225, 293–297
Binomial formula, 92–94
Binomial random variables, 86–99, 124–125
Bivariate normal distribution, 417–425, 449–450
Branching process, 68, 82

Cartesian product, 30–32, 204, 298–299
Central limit theorem, 162–168, 205, 213, 317
Centroid of bivariate distributions, 421
Chi-square distribution, 230–239, 341–343
Chi-square test:
of homogeneity, 374–380, 382–387, 395
of independence, 395–401
Cochran's Q, 464–465
Cochran's test for variances, 340–341, 346
Coefficient:
binomial, 74, 76–78
correlation, 404, 406–407, 417–418, 425–429, 433–435
phi, 401–409
Spearman's rank correlation, 455–458
Combinations, 75–79
Complementary sets, 22, 45
Composite hypotheses, 287
Conditional probability, 52–58, 266–267, 321–322, 377–380, 401–409, 493–494
Confidence intervals, 246–247, 254–255, 309, 317
for μ_X, 247–250, 252–253, 290, 292
for $\mu_1 - \mu_2$, 255–265, 312, 317
for p, 250–252, 294
for $p_1 - p_2$, 265–267, 320
for σ^2, 267–269
for σ_1^2/σ_2^2, 333–334
Contingency tables, 395–401
Continuity, correction for, 12–13, 157–158, 294–295, 390–392

Continuous random variables, 141–146
Continuous variables, 12
Contrasts, 357–360
Correlation coefficient, 404, 406–407, 417–418, 425–429, 433–435
Correlation theory, 315, 401–409, 417–438, 474–479
Counting, fundamental principle of, 68–71, 74, 100, 103
Covariance, 402–404, 424–425
  analysis of, 497–508
Covariate, 499
Cumulative probability polygon, 117–119, 143, 181

Data snooping (*see* Scheffé's method)
Decision rules, 4, 57, 91, 96, 279–281, 284–285, 298–299
Degrees of freedom, 230–232, 378, 383–384
deMorgan's laws, 26, 49
Dependent variables, 474–475
  (*See also* Independent variables)
Derived statistics, 2
Discrete random variables, 111–126
Discrete variables, 11, 111
Disjoint (mutually exclusive) sets, 23, 46–47

Efficient estimators, 202–203, 209–213
Empty sets, 21
Equal sets, 22
Events, 29
  probability of, 37–42, 248
Expected values, 120–125, 150–156, 189, 191
  (*See also* Mean value, of a population)
Experimental studies (*see* Observational studies)
Explained variability, 336–338, 349, 364–365, 401–404, 429–433, 444, 493–494

F distribution, 330–331, 341–343
F test:
  for equal mean values, 349–356
  for equal variances, 331–333
  Welch-Aspin analog to, 361–364
F^* test, 361–364
Finite sets, number of elements of, 29
Fisher's Z transformation, 449–455

Four-fold table, 53, 56–57
Frequency tables, 179–181, 186–188

Goodness of fit, 392

Hawthorne effect, 292
Histogram, 142, 181–183
Homoscedasticity, 419
Hypergeometric distribution, 132–133, 170, 225, 296–297
Hypergeometric formula, 99–104
Hypotheses:
  composite, 287
  simple, 286
Hypothesis testing theory, 94–97, 278–288, 300–303, 309, 408

Independence, statistical, 49–51, 55–58, 311, 317, 395, 401–404
Independent variables, 474–475
  (*See also* Dependent variables)
Interpolation, 183–185
Intersection of sets, 23
Interval estimation, 246–247
  (*See also* Point estimation)
Irwin-Fischer exact test, 104–105, 326, 398

K-sample test:
  of $\mu_1 = \mu_2 = \cdots = \mu_K$:
    for correlated groups, 458–464
    with equal variances, 349–361
    with unequal variances, 361–364
  of $p_1 = p_2 = \cdots = p_K$:
    for correlated groups, 464–465
    for independent groups, 377–382
  of $\rho_1 = \rho_2 = \cdots = \rho_K$, 452–455
  of $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$, 340–341

Likert scale, 139, 199, 326, 371, 443
Line graph, 51, 91, 117–119, 386–387, 400

Match pairs, 97, 263–265, 275, 315–318, 435–440, 497
  (*See also* Correlation theory)
Matching distributions, 35, 113, 121, 170, 199, 274

Mean value:
of a population, 120–125
of a sample, 185–189, 193–195
Median value:
of a population, 119
of a sample, 181, 193–195
Method of least squares, 479–488
Missing observations, 333, 341, 346
Mode:
of a population, 126
of a sample, 180
Monotonic relationships, 474–475
Monte Carlo study, 204–205
Mutually exclusive (disjoint) sets, 23, 46–47

Normal distribution, 146–150, 156–157, 162–168

Observational studies, 314–315, 317, 320–321
One-sample test:
of $H_0: \alpha = \alpha_0$, 489–490
of $H_0: \beta = \beta_0$, 489–490
of $H_0: \mu = \mu_0$:
σ^2 known, 288–291
σ^2 unknown, 291–293
of $H_0: p = p_0$, 293–297
of $H_0: \rho = \rho_0$, 450
of $H_0: \sigma^2 = \sigma_0^2$, 297–298
Orthogonal polynomials for trend, 517
Outliers, 199

Parameters, 135, 168
Pearson product moment correlation (*see* Correlation coefficient)
Percentiles, 143–144, 181–185
Permutations, 69–75
Phi coefficient, 401–409
(*See also* Correlation coefficient)
Point biserial correlation, 443–444
Point estimation, 202–203
(*See also* Interval estimation)
Populations, 6–9, 17, 261
mean value of, 120–125
median value of, 119
mode of, 126
variance of, 126–134

Post hoc comparisons (*see* Scheffé's method)
Power of a statistical test, 282, 290–291, 295–296, 301–302, 309–310, 408
Prediction, 478
Probability:
conditional, 52–58, 266–267, 321–322, 377–380, 401–409, 493–494
of an event, 37–42, 248
relative frequency interpretation of, 42–44, 47, 50, 121–122, 125, 179
Probability distribution, 51, 90, 112
Probability polygon, 143

Qualitative variables, 10
Quantitative variables, 11

Random sampling (*see* Sampling theory)
Random variables:
binomial, 86–99, 124–125
continuous, 141–146
discrete, 111–126
Regression fallacy, 439–440
Regression lines, 421–423, 439–440
Regression theory, 474–497
Relative frequency table, 180
Repeated measures, 273, 458–465
(*See also* Match pairs)

Sample:
mean value of, 185–189, 193–195
median value of, 181, 193–195
mode of, 180
variance of, 189–193, 231
Sample size determination, 269–271, 295–296, 302, 309–310
Sample space (*see* Cartesian product)
Sampling:
with replacement, 61–62, 99
without replacement, 59–61, 100–101
(*See also* Sampling theory)
Sampling distribution:
of pooled variance estimates, 338–340
of sample means, 203–207, 213–218, 225–226
of sample medians, 203–207, 213–216
of sample proportions, 223–225
of sample standard deviations, 221–223
of sample variances, 218–221, 236–239

Sampling theory, 58–62, 99–102, 174–179, 314–315, 320, 401, 439–440
Scatter diagrams, 425–428
Scheffé's method, 356–360, 371
  analog to, 364, 380–383
Sets, 19–21
  complementary, 22, 45
  disjoint (mutually exclusive), 23, 46–47
  empty, 21
  equal, 22
  finite, number of elements of, 29
  intersection of, 23
  mutually exclusive, 23, 46–47
  union of, 23, 47–49
Sigma ($\sum$), algebra of, 79–81
Sign test, 97–99, 317
Simple hypotheses, 286
Skewed distribution, 126, 194
Spearman's rank correlation coefficient, 455–458
Standard deviation (*see* Variance)
Standard errors, 251, 253–254, 261, 266, 318
Standard scores (*see* T scores)
Statistic, 179
  test, 288
Statistical independence, 49–51, 55–58, 311, 317, 395, 401–404
Statistical test·
  power of, 282, 290–291, 295–296, 301–302, 309–310, 408
  robustness of, 292, 298, 332–333, 341, 354
    (*See also* Decision rules)
Statistical variables, 9–13
Student's t distribution, 230, 239–241, 341–343
Subset, 18
Symmetric distributions, 188, 126

t distribution, 230, 239–241, 341–343
T scores, 158–162, 448–449
t test:
  one-sample, 291–293
  two-sample, 310–312
Test statistic, 288
Tied observations, 399
Tree diagram, 68, 82
Triangular distribution, 115, 130
Tukey's method, 360–361
Two-sample test:
  of $H_0: \mu_1 = \mu_2$·
    for matched pairs, 315–317
    with $\sigma_1^2 = \sigma_2^2$, but unknown, 310–312
    with unequal variances, 312–315
    with variances known, 308–310
  of $H_0: p_1 = p_2$, 318–321, 374–377
  of $H_0: \rho_1 = \rho_2$, 450–451
Type I error, 91, 281–282, 302–303, 333, 343, 360, 455, 508
Type II error, 91, 281–282, 302–303

Unbiased estimators, 202–203, 207–208
Uniform distribution, 112, 115, 128–129, 135, 162
Union of sets, 23, 47–49
Universe (*see* Populations)

Variable, 17
  Bernoulli, 86, 92, 131, 223, 293
  continuous, 12
  dependent, 474–475
  discrete, 11, 111
  independent, 474–475
  qualitative, 10
  quantitative, 11
  random (*see* Random variables)
  statistical, 9–13
Variance:
  analysis of, 349–356, 458–464, 476–479, 494–497
  Bartlett's test for, 341, 346–347
  Cochran's test for, 340–341, 346
  of correlated variables, 446–449
  of pooled samples, 257–258
  of a population, 126–134
  of a sample, 189–193, 231
Venn diagrams, 24–27

Welch-Aspin t^* distribution, 261–263, 312–315, 333
Within sample estimate of σ^2, 257–258, 334–336, 338–340